

The *CBF* gene family in hexaploid wheat and its relationship to the phylogenetic complexity of cereal *CBFs*

Mohamed Badawi · Jean Danyluk · Barbara Boucho ·
Mario Houde · Fathey Sarhan

Received: 29 September 2006 / Accepted: 23 December 2006 / Published online: 7 February 2007
© Springer-Verlag 2007

Abstract Most temperate plants tolerate both chilling and freezing temperatures whereas many species from tropical regions suffer chilling injury when exposed to temperatures slightly above freezing. Cold acclimation induces the expression of cold-regulated genes needed to protect plants against freezing stress. This induction is mediated, in part, by the CBF transcription factor family. To understand the evolution and function of this family in cereals, we identified and characterized 15 different *CBF* genes from hexaploid wheat. Our analyses reveal that wheat species, *T. aestivum* and *T. monococcum*, may contain up to 25 different *CBF* genes, and that Poaceae *CBFs* can be classified into 10 groups that share a common phylogenetic origin and similar structural characteristics. Six of these groups (IIIc, IIIId, IVa, IVb, IVc and IVd) are found only in the Pooideae suggesting they represent the CBF response machinery that evolved recently during colonization of temperate habitats. Expression studies reveal that five of the Pooideae-specific groups display

higher constitutive and low temperature inducible expression in the winter cultivar, and a diurnal regulation pattern during growth at warm temperature. The higher constitutive and inducible expression within these CBF groups is an inherited trait that may play a predominant role in the superior low temperature tolerance capacity of winter cultivars and possibly be a basis of genetic variability in freezing tolerance within the Pooideae subfamily.

Keywords Wheat · Pooideae · CBF/DREB · AP2/ERF · Evolution · Low temperature

Introduction

The ability of plants to respond and adapt to low temperature (LT) stresses varies greatly with species (Sakai and Larcher 1987). Most temperate plants, such as wheat, canola and *Arabidopsis*, are able to tolerate both chilling and freezing temperatures. In contrast, species from tropical regions, such as tomato, maize and rice are unable to tolerate freezing temperatures and even suffer chilling injury when exposed to temperatures in the range of 0–12°C. The cold acclimation process induces the expression of cold-regulated (*COR*) genes, whose products are thought to be necessary for protection against freezing stress (Thomashow 1999). Many of these *COR* genes contain copies of the C-repeat/dehydration-responsive element (CRT/DRE) in their promoters, which has the core motif CCGAC and is responsible for the LT-responsiveness of these genes. The factors that bind the CRT/DRE were first identified in *Arabidopsis* and designated CRT-Binding Factors/DRE-binding proteins 1 (*CBF/DREB1*) genes

Electronic supplementary material The online version of this article (doi:10.1007/s00438-006-0206-9) contains supplementary material, which is available to authorized users.

Communicated by J.-K. Zhu.

Mohamed Badawi and Jean Danyluk contributed equally to this work.

M. Badawi · J. Danyluk (✉) · B. Boucho · M. Houde ·
F. Sarhan
Département des Sciences biologiques,
Université du Québec à Montréal, Case Postale 8888,
Succursale Centre-ville, Montréal, QC, Canada, H3C 3P8
e-mail: danyluk.jean@uqam.ca

(Stockinger et al. 1997; Liu et al. 1998). These genes encode LT-induced transcription factors that when constitutively overexpressed in *Arabidopsis* mimic cold acclimation by inducing *COR* gene expression and freezing tolerance (FT) (Jaglo-Ottosen et al. 1998; Liu et al. 1998). CBF-like proteins have been isolated from a wide range of plants that include species capable and incapable of cold acclimation, suggesting that the CBF cold response pathway is broadly conserved in plants (Jaglo et al. 2001).

The CBF/DREB1 proteins belong to the AP2/ERF superfamily of DNA-binding proteins. This protein family has in common the AP2 DNA-binding motif. The 122 *Arabidopsis* members of the ERF family containing one AP2 DNA-binding motif were divided into 12 groups, with several of the groups being further divided into subgroups (Nakano et al. 2006). The six CBF/DREB1 proteins of *Arabidopsis* were included in subgroup IIIc, one of the five group III subgroups. The CBF proteins are distinguished from other group III members by a conserved set of amino acid sequences (motif CMIII-3) flanking the AP2 DNA-binding domain. Other motifs found in CBF proteins (CMIII-1, CMIII-2 and CMIII-4) are also present in one or more of the group III subgroups, suggesting related molecular functions may be conserved between subgroups. These features were previously noted as conserved features of the C-terminal activation domain between aligned CBFs (Wang et al. 2005; Skinner et al. 2005). Comparison of group III proteins between the eudicot *Arabidopsis* and the monocot rice reveals they share four common subgroups (Nakano et al. 2006) suggesting a functional diversification of group III proteins had occurred before the divergence of these two species. The 4 ancestral genes have amplified to 23 and 26 genes in *Arabidopsis* and rice, respectively. *Arabidopsis* CBF studies have demonstrated that *CBF1*, 2, and 3 function in the cold-acclimation pathway with redundant and some possibly specific functions (Gilmour et al. 2004; Novillo et al. 2004; Van Buskirk and Thomashow 2006), *CBF4* is involved in drought adaptation (Haake et al. 2002), and *DDF1* and *DDF2* are involved in gibberellin biosynthesis and salt stress tolerance (Magome et al. 2004). Based on these findings, the functional divergence of group III ERF proteins in monocots may differ from eudicots and therefore warrant a characterization of monocot genes.

Members of the Poaceae have been targeted for study since they contain the major cereal crops wheat, maize and rice which provide >60% of the calories and proteins for our daily life. To meet the needs of the projected human population by 2050, cereal grain production must increase at an annual rate of 2% on an

area of land that will not increase much beyond the present level (Gill et al. 2004). Therefore, significant advances in our understanding of the *CBF* family in cereals are essential to develop needed strategies to protect crops from losses caused by abiotic stress. The Poaceae represent an excellent model system to study the roles of the *CBF* family in the evolution of LT tolerance. The Poaceae radiated some 55–70 million years ago (MYA) into several subfamilies (Kellogg 2001). The subfamilies Oryzaceae (rice) and Panicoideae (maize) have a more tropical geographical distribution compared to members of the Pooideae, which contain the temperate cereals wheat, barley and oat. LT tolerance within the Pooideae subfamily ranges from low in oat (a Poaceae tribe representative), to intermediary in barley and wheat (Triticeae tribe), to highly tolerant in rye (Triticeae tribe). The estimated divergence time between the Triticeae and Poaceae is around 35 MYA, and within the Triticeae, barley and rye diverged from wheat around 11 and 7 MYA, respectively (Huang et al. 2002a). The more recent evolutionary history of bread wheat started with an adaptive radiation of the diploid progenitors around 2.2–4.5 MYA followed by successive hybridizations around <0.5 MYA and 8,000 years ago to produce hexaploid bread wheat (Huang et al. 2002b). Wheat is an interesting model since the comparative analysis of *CBF* gene function among closer and more distantly related species may shed light on important evolutionary trends that have sculpted *CBF* function. Furthermore, the three genomes of hexaploid wheat are known to contain differences for many agronomically important genes (Gill et al. 2004), and recently, a LT tolerance QTL on chromosome 5 of *T. monococcum* (Miller et al. 2006), barley (Skinner et al. 2006) and hexaploid wheat (Båga et al. 2007) was found to coincide with the location of 11, 12 and 2 *CBF* genes, respectively. The exact molecular explanation for this LT tolerance QTL is not known but indicates the possibility that *CBF* genes may be at the base of this important trait.

Many *CBF* genes have been identified from Poaceae species such as rye (Jaglo et al. 2001), rice (Dubouzet et al. 2003; Skinner et al. 2005), barley (Choi et al. 2002; Xue 2002; Francia et al. 2004; Skinner et al. 2005), wheat (Jaglo et al. 2001; Kume et al. 2005; Skinner et al. 2005; Vágújfalvi et al. 2005; Miller et al. 2006), and *Festuca arundinacea* (Tang et al. 2005). These studies, in particular those of Skinner et al. (2005) and Miller et al. (2006), have revealed that the cereal *CBF* family is large and complex. To better understand the functions of this gene family during cold stress and its evolution in the Poaceae, we initiated a study to identify and characterize *CBF* genes

from hexaploid wheat. Here, we show that hexaploid wheat contains at least 15 different *CBF* genes, and that Poaceae *CBFs* can be subdivided into groups with specific characteristics. These findings expand our understanding on the functional categories of cereal *CBF* genes and provide a starting point for future studies.

Materials and methods

Preparation of the cDNA libraries

Five different cDNA libraries prepared from *Triticum aestivum* L. cv Norstar were used to identify expressed wheat *CBF* genes in this study (Table S1). Plant growth conditions, RNA purification and cDNA library construction were described in detail elsewhere (Houde et al. 2006). Briefly the five libraries (L2–L6) were prepared from the following pooled mRNA populations: (L2) aerial parts (leaf and crown) from control and long-term cold acclimated wheat (1–53 days); (L3) root tissue from control, cold-acclimated and salt stressed wheat; (L4) aerial parts of dehydration stressed wheat; (L5) crown tissue during vernalization and different developmental stages of spike and seed formation in wheat; (L6) crown and leaf tissues from wheat after short exposures to LT in the light and in the dark. All cDNAs synthesized were directionally cloned into the pCMV.SPORT6 vector (Invitrogen) with the *SalI* adaptor (GTCGACCCACGCGTCCG) and *NotI* primer adaptor (GCGGCCGCCCT₁₅). For the last four libraries, the first strand cDNA reaction mix contained methylated dCTP to prevent cDNAs from internal cleavage by the *NotI* restriction enzyme used for directional cloning. For the last library, the ‘GeneRacer’ kit (Invitrogen) was used prior to first strand synthesis to produce a library containing a high proportion (95%) of full-length cDNAs. For each library, six million primary transformants were obtained, amplified and frozen as glycerol stocks.

To prepare plasmids for PCR experiments, 40 µl of a bacterial library stock ($>40 \times 10^6$ clones) were inoculated into LB media (100 ml) supplemented with ampicillin and grown at 30°C for 6–8 h. Plasmids were isolated using the QIAprep Miniprep system (QIAGEN) and the quantity was evaluated on a gel.

Identification of wheat *CBF* genes

The gene cloning approach, gene names and GenBank accession numbers are summarized in Table 1. To initiate this project, available *CBF* protein sequences were used for data mining of the NCBI NR and EST

databases in search of wheat homologs. The mRNA (Jaglo et al. 2001) and EST sequences were assembled into virtual mRNAs using CAP3 (<http://www.pbil.univ-lyon1.fr/cap3.php>) (Huang and Madan 1999). This initial assembly was updated as additional *CBF* genes were sequenced from the project. Virtual mRNAs that contained EST sequences from the wheat genomics of abiotic stress (WGAS) project (<http://www.bioinfouqam.wgas.ca/cgi-bin/abiotic/project.cgi>) were ordered and completely sequenced at the Génome Québec sequencing center (McGill University, Canada). Initially, hybridization probes corresponding to either incomplete *CBF* genes or to their AP2 DNA binding region were used to screen the wheat plasmid cDNA libraries L3 and L6 to identify additional full length *CBF* clones. Because this approach was time consuming, a PCR strategy was used to identify *CBF* genes expressed in wheat. From one to three gene specific primers (GSPs) were designed using Primer3 (http://www.frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) (Rozen and Skaletsky 2000) in the forward (5′ mRNA region) or reverse (3′ mRNA region) directions on the assembled virtual mRNAs. In order to drive the PCR reaction, the GSPs were designed with a higher T_m than the universal primers M13F 5′-AGATCCCAAGCTAGCAG TTTTCCAGTCACGA-3′ and M13R 5′-GAGCGG ATAACAATTTACACAGGAAACAGCTATGA-3′ found in the pCMV.SPORT6 vector. A list of GSPs used in the isolation of *CBF* genes is given in Table S2. Amplification was performed using a GSP (0.6 µM) and the corresponding universal primer (0.3 µM), 10 ng of plasmid DNA from a library, 1× or 2× enhancer solution, and *Pfx* DNA polymerase (Invitrogen) following the manufacturer’s recommendations. Briefly, the PCR thermal-cycling parameters were initiated with a progressive step down in annealing temperature that ended with 45 cycles of 94°C for 20 s, 64°C for 20 s and 68°C for 150 s. PCR products were analyzed on a gel and lanes that produced DNA fragments of expected size were chosen for subcloning using the Zero Blunt TOPO PCR Cloning Kit (Invitrogen) and then sent for sequencing. Novel *CBF* sequences were fully sequenced and overlapping sequences were merged to produce the longest possible gene sequence (Table 1). Several clones from independent amplification reactions were sequenced in order to correct errors introduced during PCR.

Expression profiling using northern analysis

The spring wheat cultivar *Triticum aestivum* L. cv Quantum and the winter cultivar Norstar were germinated in moist sterilized vermiculite in a growth

Table 1 Nomenclature and characteristics of wheat *CBF* genes isolated from the cultivar Norstar

Gene name	Accession number	Isolation method ^a	Chromosome location ^b	cDNA (bp) ^c	Open reading frame (CDS)	Total amino acids
<i>TaCBF1a-A11</i>	EF028751	3	2AL	962	54–710	218
<i>TaCBF11-5.1</i>	EF028752	3		1,069	98–775	225
<i>TaCBF11-5.2</i> ^d	EF028753	3		991	109–768	219
<i>TaCBF11-5.3</i>	EF028754	3		876	107–793	228
<i>TaCBF11a-6.1</i>	EF028755	1		958	85–795	236
<i>TaCBF11a-6.2</i>	EF028756	3		947	112–840	242
<i>TaCBF11a-D6</i>	EF028757	2	5DL-1; 5DL-2	959	107–823	238
<i>TaCBF11c-3.1</i>	EF028758	3		896	<1–708	>235
<i>TaCBF11c-3.2</i>	EF028759	3		922	111–851	246
<i>TaCBF11c-D3</i>	EF028760	1	5DL-1; 5DL-2	963	105–842	245
<i>TaCBF11c-B10</i>	EF028761	1, 3	5B	1,000	91–813	240
<i>TaCBF11d-12.1</i>	EF028762	3		1,033	130–867	245
<i>TaCBF11d-B12</i>	EF028763	1, 3	5B	970	88–825	245
<i>TaCBF11d-A15</i>	EF028764	3	5AL-10; 5AL-8	1,010	107–826	239
<i>TaCBF11d-15.2</i>	EF028765	3		926	110–835	241
<i>TaCBF11d-A19</i>	EF028766	2, 3	5AL-10; 5AL-8	984	184–888	234
<i>TaCBF11d-B19</i>	EF028767	1	5B	954	150–854	234
<i>TaCBF11d-D19</i>	EF028768	1	5DL-1; 5DL-2	965	166–870	234
<i>TaCBF11a-A2</i>	EF028769	2, 3	5AL-10; 5AL-8	950	52–729	225
<i>TaCBF11a-2.2</i> ^e	EF028770	3		822	<3–693	>230
<i>TaCBF11a-2.3</i>	EF028771	3		660	86–>660	>191
<i>TaCBF11b-A20</i>	EF028772	3	5AL-10; 5AL-8	1,077	77–730	217
<i>TaCBF11b-B20</i>	EF028773	1	5B	887	63–701	212
<i>TaCBF11b-D20</i>	EF028774	2, 3	5DL-1; 5DL-2	980	77–715	212
<i>TaCBF11b-21.1</i>	EF028775	3		979	193–801	202
<i>TaCBF11b-D21</i>	EF028776	3	5DL-1; 5DL-2	1,066	380–988	202
<i>TaCBF11c-14.1</i>	EF028777	1, 3		1,003	120–758	212
<i>TaCBF11c-B14</i>	EF028778	2	5B	863	106–750	214
<i>TaCBF11c-14.3</i>	EF028779	1		892	119–763	214
<i>TaCBF11d-4.1</i>	EF028780	1		872	53–721	222
<i>TaCBF11d-B4</i>	EF028781	1	5B	865	56–724	222
<i>TaCBF11d-9.1</i>	EF028782	2		1,080	99–908	269
<i>TaCBF11d-B9</i>	EF028783	2	5B	999	77–886	269
<i>TaCBF11d-D9</i>	EF028784	2, 3	5DL-1; 5DL-2	1,063	110–919	269
<i>TaCBF11d-A22</i>	EF028785	3	5AL-10; 5AL-8	1,219	95–922	275
<i>TaCBF11d-B22</i> ^f	EF028786	1, 3	5B	1,252	94–966	290
<i>TaCBF11d-D22</i>	EF028787	3	5DL-1; 5DL-2	1,211	96–923	275

^a The methods used to isolate *CBF* genes in wheat are described by the codes: 1 WGAS EST; 2 cDNA library screening; 3 PCR amplification from cDNA libraries

^b Indicates either a location to a chromosome, a chromosome arm or a deletion breakpoint interval

^c The total number of bases in a cDNA does not include the poly A tail and gene specific primer sequences if they were not confirmed by additional sequence information

^d Mapping did not produce a chromosome location. However, orthologs in *T. monococcum* (Miller et al. 2006) and barley (Skinner et al. 2006) were mapped to chromosomes 7A and short arms of 7H, respectively

^e Contains a T→C transition that destroys a termination codon and extends the coding sequence for 93 bases

^f Contains a 38 base insertion that creates a frame shift for the remaining coding sequence

chamber (Model-E15, Conviron) for 7 days at 20°C and 70% relative humidity under an irradiance of 250 μmol m⁻² s⁻¹ and a 16-h photoperiod. At the end of this period, 5 g of control leaves were sampled and individually frozen. A cold treatment (4°C) was initiated by changing the temperature in the growth chamber 4 h into the day phase and continued under the same irradiance and photoperiod conditions for the times indicated in Fig. 4. Total RNA was extracted

from wheat leaves as described (Danyluk and Sarhan 1990), and equal amounts (10 μg) were separated on formaldehyde–agarose gels. Transfers to positively charged nylon membranes and hybridizations with ³²P-labeled probes were performed using standard molecular biology techniques. Probes for the different *TaCBFs* were designed outside of the AP2 domain to avoid cross-hybridization with known CBF transcripts (Table S3). All filters were washed at high stringency

($0.1\times$ SSC, 0.1% SDS; 65°C) for 30 min. Membranes were exposed to BioMax-MS films (Kodak) and Molecular Imager FX screens (Bio-Rad).

Expression profiling by quantitative real-time PCR

Plant growth conditions and cDNA synthesis

Spring cultivar Manitou and winter cultivar Norstar were germinated in a mixture of 50% black earth and 50% Pro-Mix (Premier) for 8 days at 20°C under a 16-h photoperiod with a light intensity of $250\ \mu\text{mol m}^{-2}\ \text{s}^{-1}$. In this experiment, Manitou was used as the spring cultivar in order to compare the response to the Quantum spring cultivar. At the end of this period, 8 control seedlings were sampled every 2 h on dry ice, and immediately frozen at -70°C . Cold treatment (4°C) was initiated by changing the temperature in the growth chamber 4 h before the night cycle, and sampling as indicated in each Figure. Sampling of control and cold-treated plants every 2 h was chosen so that closely spaced time points act as closely related biological replicates in addition to providing a better resolution of the expression characteristics of a gene. Night samples were collected by opening the growth chamber in the dark, removing plants to another room and harvesting rapidly. Total RNA was isolated using the RNeasy Plant Mini Kit (QIAGEN) using the optional on-column DNase digestion. Purified RNA ($2.8\ \mu\text{g}$) was reverse transcribed in a $20\ \mu\text{l}$ reaction volume using the Superscript II first strand cDNA synthesis system for RT-PCR (Invitrogen). Parallel reactions for each RNA sample were run in the absence of Superscript II (no RT control) to assess for genomic DNA contamination. The reaction was terminated by heat inactivation; the cDNA product was treated with RNase H, and diluted in water ($20\ \text{ng}/\mu\text{l}$) for storage (-20°C).

Design of gene-specific primers

The genome of hexaploid wheat contains three genomes inherited from three diploid ancestors. The 37 *TaCBF* gene sequences identified in this study were analyzed using ClustalW (<http://www.align.genome.jp/>) and phylogenetic characterization. This analysis revealed 15 groups of genes containing one to three homeologous copies in each group. Primers were specifically designed to monitor the expression of only one representative gene per group. This representative was chosen randomly. Fluorescent TaqMan-MGB probes as well as the non-fluorescent primers (Table S4) were designed using the combination of Primer Express Software Version 2.0 (Applied Biosystems) and

Primer3. BLASTN searches against EST and NR databases were performed to confirm the gene specificity of the primers. Non-fluorescent primers were synthesized by Invitrogen and TaqMan-MGB probes by Applied Biosystems.

PCR amplification and data analysis

Quantitative real-time PCR assays for each gene target were performed in quadruplicate on an ABI Prism 7000 sequence detection system (Applied Biosystems) using the eukaryotic 18S rRNA as the endogenous control (Applied Biosystems #4319413E). From the diluted cDNA, $2\ \mu\text{l}$ ($40\ \text{ng}$) was used as a template in a $25\text{-}\mu\text{l}$ PCR reaction containing $1\times$ TaqMan universal PCR master mix (Invitrogen), $0.9\ \mu\text{M}$ of non-fluorescent primers, and $0.25\ \mu\text{M}$ of TaqMan-MGB fluorescent probe. The PCR thermal cycling parameters were 50°C for 2 min, 95°C for 10 min followed by 50 cycles of 95°C for 15 s and 60°C for 1 min.

All calculations and statistical analysis were performed by the SDS RQ Manager 1.1 software using the $2^{-\Delta\Delta\text{Ct}}$ method with a relative quantification $\text{RQ}_{\text{min}}/\text{RQ}_{\text{max}}$ confidence set at 95% (Livak and Schmittgen 2001). The error bars display the calculated maximum (RQ_{max}) and minimum (RQ_{min}) expression levels that represent SE of the mean expression level (RQ value). Collectively, the upper and lower limits define the region of expression within which the true expression level value is likely to occur (SDS RQ Manager 1.1 software user manual; Applied Biosystems). Amplification efficiency (90–100%) for the 15 primer sets was determined by amplification of cDNA dilution series using 80, 20, 10, 5, 2.5, and 1.25 ng per reaction (data not shown). Specificity of the RT-PCR products was assessed by gel electrophoresis.

Chromosome localization of *TaCBF* genes

Genomic DNA was extracted and quantified (Limin et al. 1997) from several stocks of the wheat cultivar Chinese Spring: ditelocentric series provided by the USDA from E. R. Sears collection (all chromosomes are present but in each line one chromosome pair is represented by only the telocentric chromosomes of one arm); chromosome 5 nullisomic–tetrasomic lines (a pair of chromosomes is removed and replaced by another pair of homoeologous chromosomes); and deletion lines for homoeologous group 5AL and 5DL chromosomes (Endo 1988; Endo and Gill 1996) generated using the gametocidal chromosome of *Aegilops cylindrical*. From the diluted genomic stocks, $2\ \mu\text{l}$ ($20\ \text{ng}$) was used as a template in a $25\text{-}\mu\text{l}$

PCR reaction containing 1× TaqMan universal PCR master mix (Invitrogen), 0.9 μM non-fluorescent primers, and 0.25 μM TaqMan-MGB fluorescent probe. The PCR thermal cycling parameters were 50°C for 2 min, 95°C for 10 min followed by 50 cycles of 95°C for 15 s and 60°C for 1 min. At the end of the run, the Ct values were compared, and genetic stocks that showed a delayed or undetectable amplification were identified as the location of the assayed *TaCBF* gene.

Phylogenetic and other bioinformatic analysis

Monocotyledon *CBF* homologs were identified using the *TaCBF* nucleotide and protein sequences as queries against the GenBank NR and EST databases. Overlapping ESTs were assembled into virtual cDNAs using CAP3 (<http://www.pbil.univ-lyon1.fr/cap3.php>) (Huang and Madan 1999) and a consensus cDNA sequence was deduced. EST-derived sequences for analyses were obtained by trimming edges that corresponded to low quality error prone regions which were revealed through blastx searches against the NR database. Accession numbers for genes and ESTs used in this study are in Table 2. FASTA files of nucleotide and protein sequences used in this analysis are presented in Supplemental Tables S5 and S6. The degree of sequence identity was determined using ALIGN and FASTA on the Biology Workbench (<http://www.workbench.sdsc.edu>). Sequences were aligned using ClustalW from the Biology Workbench or from the MEGA software version 3.1 (Kumar et al. 2004), and alignments were refined manually. The MEGA software was used for phylogenetic analyses and the Minimum Evolution tree was derived from this alignment using the Kimura 2-parameter with bootstrap test and default parameters.

Hydrophobic cluster analysis (HCA) (Gaboriaud et al. 1987; Callebaut et al. 1997) was conducted using the web-based interface at: (http://www.bioserv.rpbs.jussieu.fr/RPBS/cgi-bin/Ressource.cgi?chzn_lg = an&chzn_rsrc = HCA). Briefly, the protein sequences are displayed on a duplicated α -helical net in which hydrophobic amino acids (V, I, L, F, M, Y, W) are contoured. Hydrophobic residues separated by four or more nonhydrophobic residues, or a Proline, are placed into distinct clusters. The defined hydrophobic clusters were shown to mainly correspond to the internal faces of regular secondary structures (α -helices or β -strands). Two other amino acids were chosen to be highlighted in this study: proline which confers the greatest constraint to the polypeptide chain and glycine which confers the largest freedom to the chain. This

secondary structure information was highlighted on a ClustalW alignment of group-related wheat *CBFs*.

Results

Identification of wheat *CBF* genes

CBF family members are important regulators of FT in plants. Data mining and analyses of cereal *CBF* sequences present in GenBank suggest that different species contain diverse and complex *CBF* families. Based on preliminary studies conducted in a number of varieties, hexaploid wheat contains at least seven *CBF* genes (Jaglo et al. 2001; Kume et al. 2005; Skinner et al. 2005). To maximize our chance of discovering *CBF* genes involved in the development of wheat FT, several cDNA libraries were constructed (Table S1) and screened to identify *CBF* genes expressed under various cold acclimation time points and conditions in the freezing tolerant cultivar Norstar. A combination of EST sequencing, cDNA library screening and PCR amplification allowed the identification of 37 expressed *TaCBF* genes from hexaploid wheat (Table 1). To be consistent with the established *H. vulgare* and *T. monococcum* nomenclature (Skinner et al. 2005; Miller et al. 2006), we assigned identical gene numbers to orthologs of hexaploid wheat (2–15) and new consecutive numbers (starting from 19 to 22) to novel genes identified following homology comparison with published and identified wheat sequences (e.g. *TaCBF1a-A11* shows the highest homology with its ortholog *HvCBF11* identified previously (Skinner et al. 2005)). These analyses revealed that the 37 genes identified from hexaploid wheat can be classified into at least 15 different orthologous gene groups with 1–3 homeologous copies in each. Phylogenetic analysis of *T. aestivum* and *T. monococcum* genes reveals that wheat *CBF* genes can be divided into 10 monophyletic groups. Therefore, we included in the proposed *CBF* nomenclature *CBF* subgroup information (e.g. *CBF1a*, II, IIIa, IIIb, IIIc, IIId, IVa, IVb, IVc and IVd) which should facilitate future comparison of monocot *CBF* properties and functions. In the cases where homeologous copies were mapped to one of the three genomes of hexaploid wheat, a letter designating its location precedes the *CBF* gene number (e.g. *TaCBFIIIa-D6*). On the other hand, when the genomic localization of a homeologous copy has not yet been determined, the *CBF* gene number is followed by a temporary designation of .1, .2 or .3 (e.g. *TaCBFIIIa-6.1*).

At least three gene groups (*TaCBFIIIId-19*, *TaCBFIVb-20* and *TaCBFIVd-22*) represent true

Table 2 List of monocotyledon *CBF* genes and their proposed nomenclature

Plant species ^a	Proposed gene name	Gene name	Accession number	Type of sequence
Poales/Poaceae/Oryzaceae				
<i>Oryza sativa</i>	<i>OsCBFI-1F</i>	<i>OsDREB1F</i>	AY785897	mRNA
	<i>OsCBFIa-1G</i>	<i>OsDREB1G</i>	AK060550	mRNA
	<i>OsCBFIa-1E</i>	<i>OsDREB1E</i>	AY785896	mRNA
	<i>OsCBFII-1C</i>	<i>OsDREB1C</i>	AY327040	mRNA
	<i>OsCBFIII-1D</i>	<i>OsDREB1D</i>	AY785895	mRNA
	<i>OsCBFIII-1I</i>	<i>OsDREB1I</i>	XM_483622	Genomic
	<i>OsCBFIII-1J</i>	<i>OsDREB1J</i>	XM_483621	Genomic
	<i>OsCBFIIIa-1A</i>	<i>OsDREB1A</i>	AF300970	mRNA
	<i>OsCBFIIIb-1H</i>	<i>OsDREB1H</i>	AP008215	Genomic
	<i>OsCBFIV-1B.1</i>	<i>OsDREB1B.1</i>	(CDS: 20099705-20098965) AY785894	mRNA
Poales/Poaceae/Pooideae				
<i>Agrostis capillaris</i>	<i>AcCBFIIIa-6</i>		DV858477	EST
	<i>AcCBFIIIc-19</i>		DV853050	EST ^b
<i>Agrostis stolonifera</i>	<i>AstCBFIIIc-3</i>		DY543542	EST ^b
<i>Avena sativa</i>	<i>AsCBFIIIc-12</i>	<i>CBF4</i>	AM071409	mRNA
	<i>AsCBFIIIc-16A</i>	<i>CBF1</i>	AM071406	mRNA
	<i>AsCBFIIIc-16B</i>	<i>CBF2</i>	AM071407	mRNA
	<i>AsCBFIVa</i>		AM071408	EST ^b
<i>Brachypodium distachyon</i>	<i>BdCBFIa-1</i>		DV479443	EST
	<i>BdCBFII-5</i>		DV482761, DV481106, DV485162	EST
	<i>BdCBFIIIa-6</i>		DV485858, DV489297, DV487222, DV478843, DV489355, DV484024, DV478031	EST
<i>Festuca arundinacea</i>	<i>FaCBFIIIa-6</i>	<i>DREB1A</i>	AJ717399	mRNA
	<i>FaCBFIIIc-3</i>		DT710563, DT701428	EST
	<i>FaCBFIVa-2</i>	<i>DREB1</i>	AY423713	mRNA
<i>Hordeum brevisubulatum</i>	<i>HbCBFIVa-2</i>	<i>DREB1</i>	DQ250027	mRNA
<i>Hordeum vulgare</i>	<i>HvCBFIa-1</i>	<i>HvCBF1</i>	AY785836	Genomic
	<i>HvCBFIa-11</i>	<i>HvCBF11</i>	AY785890	Genomic
	<i>HvCBFII-5</i>	<i>HvCBF5</i>	AY785855	Genomic
	<i>HvCBFIIIa-6</i>	<i>HvCBF6</i>	AY785860	Genomic
	<i>HvCBFIIIc-3</i>	<i>HvCBF3</i>	AY785845	Genomic
	<i>HvCBFIIIc-8A</i>	<i>HvCBF8A</i>	AY785868	Genomic
	<i>HvCBFIIIc-10A</i>	<i>HvCBF10A</i>	AY785882	Genomic
	<i>HvCBFIIIc-13</i>	<i>HvCBF13</i>	DQ095158	Genomic
	<i>HvCBFIIIc-12</i>	<i>HvCBF12</i>	DQ095157	Genomic
	<i>HvCBFIVa-2A</i>	<i>HvCBF2A</i>	AY785841	Genomic
	<i>HvCBFIVc-14</i>	<i>HvCBF14</i>	DQ095159	Genomic
	<i>HvCBFIVd-4A</i>	<i>HvCBF4A</i>	AY785849	mRNA
	<i>HvCBFIVd-9</i>	<i>HvCBF9</i>	AY785878	Genomic
<i>Lolium perenne</i>	<i>LpCBFIIIa-6</i>	<i>CBF3</i>	AY960831	mRNA
<i>Secale cereale</i>	<i>ScCBFIVb-20</i>	<i>Clone 79 CBF</i>	AF370728	mRNA
	<i>ScCBFIVd-9A</i>	<i>Clone 80 CBF</i>	AF370729	mRNA
	<i>ScCBFIVd-9B</i>	<i>Clone 81 CBF</i>	AF370730	mRNA
<i>Triticum aestivum</i>	<i>TaCBFIa-A11</i>		EF028751	mRNA
	<i>TaCBFII-5.1</i>		EF028752	mRNA
	<i>TaCBFIIIa-6.1</i>		EF028755	mRNA
	<i>TaCBFIIIc-3.1</i>		EF028758	mRNA
	<i>TaCBFIIIc-B10</i>		EF028761	mRNA
	<i>TaCBFIIIc-12.1</i>		EF028762	mRNA
	<i>TaCBFIIIc-A15</i>		EF028764	mRNA
	<i>TaCBFIIIc-A19</i>		EF028766	mRNA
	<i>TaCBFIVa-A2</i>		EF028769	mRNA
	<i>TaCBFIVb-A20</i>		EF028772	mRNA
	<i>TaCBFIVb-21.1</i>		EF028775	mRNA
	<i>TaCBFIVc-14.1</i>		EF028777	mRNA
	<i>TaCBFIVd-4.1</i>		EF028780	mRNA
	<i>TaCBFIVd-9.1</i>		EF028782	mRNA
	<i>TaCBFIVd-A22</i>		EF028785	mRNA

Table 2 continued

Plant species ^a	Proposed gene name	Gene name	Accession number	Type of sequence
<i>Triticum Monococcum</i>	<i>TmCBFII-5</i>	<i>TmCBF5</i>	AY951947	Genomic
	<i>TmCBFIIIb-18</i>	<i>TmCBF18</i>	AY951946	Genomic
	<i>TmCBFIIIc-10</i>	<i>TmCBF10</i>	AY951950	Genomic
	<i>TmCBFIIIc-13</i>	<i>TmCBF13</i>	AY951951	Genomic
	<i>TmCBFIII d-16</i>	<i>TmCBF16</i>	AY951944 (CDS: 181665-182366)	Genomic
	<i>TmCBFIII d-17</i>	<i>TmCBF17</i>	AY951945 (CDS: 91223-90357)	Genomic
	<i>TmCBFIVa-2</i>	<i>TmCBF2</i>	AY951945 (CDS: 22116-22838)	Genomic
	<i>TmCBFIVd-4</i>	<i>TmCBF4</i>	AY951945 (CDS: 35360-34722)	Genomic
Poales/Poaceae/Panicoideae				
<i>Panicum virgatum</i>	<i>PvCBFIa-11</i>		DN144490, DN145877	EST
	<i>PvCBFII-5</i>		DN143696, DN145297, DN143313	EST
	<i>PvCBFIIIa-6</i>		DN144355, DN143145, DN143526	EST
<i>Saccharum officinarum</i>	<i>SoCBFIa-11</i>		CA212714, CA156028, CA080013, BQ534420	EST
	<i>SoCBFII-5</i>		CA162524, CA161788, CA089833, BQ533805	EST
	<i>SoCBFIIIa-6</i>		BU103690, CA090085, CA109731, CA105526	EST
	<i>SoCBFIIIb</i>		CA155733	EST ^b
	<i>SoCBFIV</i>		CA274081	EST ^b
<i>Sorghum bicolor</i>	<i>SbCBFII-5</i>	<i>SbCBF5</i>	AY785898	mRNA
	<i>SbCBFIIIa-6</i>	<i>SbCBF6</i>	AY785899	mRNA
<i>Zea mays</i>	<i>ZmCBFII-5</i>		DR964417, DV523865, DR790548, DV530932	EST
	<i>ZmCBFIIIb-1A</i>	<i>DREB1A</i>	AF450481	mRNA
	<i>ZmCBFIIIb-1B</i>		EB674710, DV539998, DV510250, DR821363, EB674709, EB400670, DR795602	EST
Arecales/Arecaceae/Coryphoideae				
<i>Sabal palmetto</i>	<i>SpCBFI-1</i>	<i>CBF-like</i>	DQ497730	Genomic
<i>Rhapidophyllum hystrix</i>	<i>RhCBFI-1</i>	<i>CBF-like 1</i>	DQ497742	mRNA
	<i>RhCBFI-2</i>	<i>CBF-like 2</i>	DQ497743	mRNA
Arecales/Arecaceae/Arecoideae				
<i>Dypsis lutescens</i>	<i>DICBFI-1</i>	<i>CBF-like</i>	DQ497738	Genomic
Zingiberales/Zingiberaceae				
<i>Zingiber officinale</i>	<i>ZoCBFI-1</i>		DY354914, DY380878	EST
	<i>ZoCBFI-2</i>		DY344903, DY345420, DY380161, DY381223	EST

^a Plant species are organized under the appropriate monocot order/family/subfamily

^b The proposed gene names are tentative since they are based on partial EST sequences

orthologous series in hexaploid wheat since the homeologous copies (A, B and D) were identified and mapped to each genome equivalent in our study (Table 1). One gene group (*TaCBFII-5*) was not mapped in our study (Table 1). However, the *T. monococcum* ortholog was mapped to chromosome 7A (Miller et al. 2006) and the barley ortholog was mapped to the short arm of 7H (Skinner et al. 2006) suggesting that this gene group will be located on chromosome 7. The vast majority of gene groups (13 out of 15) were mapped to chromosome 5 (Table 1), and at least 9 of these 13 have so far been mapped more precisely to a region, between two deletion breakpoints, associated with a cold tolerance QTL in several Triticeae species (Francia et al. 2004; Båga et al. 2007; Miller et al. 2006; Skinner et al. 2006). The present

study allowed the localization of 4 new gene groups (*TaCBFIII d-19*, *TaCBFIVb-20*, *TaCBFIVb-21* and *TaCBFIVd-22*) (Table 1) to this QTL region containing the 11 tandem *CBF* genes in *T. monococcum* (Miller et al. 2006) and 12 tandem *CBF* genes in barley (Skinner et al. 2006).

A cut-off of 95% was chosen to differentiate between homeologous copies and possible recently duplicated genes. Because homeologous copies of different *CBF* gene groups would have started diverging around the same time, one would expect them to show similarities in the same range. The comparison of *T. monococcum* (Miller et al. 2006) and *T. aestivum* *CBF* genes (this study) does reveal five orthologous genes with high levels of identity (more than 98%). From the 13 *T. aestivum* gene groups that contain

homeologous copies, only 5 show identities below 95% (the percent identity excludes bases involved in transitions, transversions and gaps) between the homeologous ORFs. These include *TaCBFII-5*, *TaCBFIIIa-6*, *TaCBFIIIc-3*, *TaCBFIVd-9* and *TaCBFIVd-22* which show identities of 85.9, 91.7, 94.5, 94.8 and 89.9%, respectively. When the identity comparison is repeated with gap regions not included, only *TaCBFII-5* and *TaCBFIIIa-6* show lower similarities of 93.4 and 93.5%, respectively. Therefore, these results suggest that the *TaCBFII-5* and *TaCBFIIIa-6* homeologous groups may contain closely related paralogs and/or have diverged at a different rate than other groups.

Phylogeny of monocot *CBF* genes

Monocot and eudicot *CBF* sequences are separated on a phylogenetic tree (Dubouzet et al. 2003; Qin et al. 2004; Bräutigam et al. 2005; Xiong and Fei 2006) suggesting that at least some *CBF* gene function specialization has evolved recently in plants. To understand the relationship between *TaCBFs* and other monocot *CBFs*, we searched NR and EST databases and compiled a set of sequences for analysis (Table 2). During our BLAST comparisons of *CBFs*, we found that it would be difficult to align the entire ORF of distant members with complete confidence. Therefore, the nucleotide sequence of the AP2 DNA binding domain and adjacent *CBF* signatures were chosen for alignment and phylogeny analysis since these domains are extremely well conserved in the *CBF* family. To simplify the future comparison of *CBF* gene functional studies with different monocotyledon plants, we propose a nomenclature that reflects the evolutionary relationship of *CBF* genes. This will help to distinguish their specific functional roles which may have appeared during their evolution.

To establish the relationship between the different *CBF* genes in wheat, only one representative of each of the 15 *TaCBF* gene groups (preferentially the homeologous A copy) was included in this analysis. In addition, 8 (Table 2) of the 13 identified *T. monococcum* *CBF* genes (Miller et al. 2006) were included in the analysis since they showed less than 95% identity (gap regions not included in the analysis) with any of the 15 *TaCBF* gene groups. These lower identities of *TmCBFII-5* (86.5%), *TmCBFIIIb-18* (76%), *TmCBFIIIc-10* (94%), *TmCBFIIIc-13* (86.7%), *TmCBFIIIId-16* (82.1%), *TmCBFIIIId-17* (78.8%), *TmCBFIVa-2* (92.4%) and *TmCBFIVd-4* (87.1%) with their closest homologues suggest that they may represent additional wheat *CBF* genes. To compare wheat *CBFs* with the closely related Triticeae species barley,

we included 13 of the 19 *HvCBF* genes reported (Table 2) (Skinner et al. 2005). The two additional pseudogenes, *HvCBFIIIc-8B* and *HvCBFIIIc-8C*, and the remaining four genes *HvCBFIIIc-10B* (97.2%), *HvCBFIVa-2B* (98.8%), *HvCBFIVd-4B* (99.8%) and *HvCBFIVd-4D* (96.4%) were not included since their high homologies with their closest related paralog suggested that these duplications happened following the divergence of wheat and barley. Several other *CBF* sequences from families of the monocotyledon order Poales were included in the analysis (Table 2). In addition, *CBF* representatives of two other monocotyledon orders Arecales and Zingiberales were also included in the analysis (Table 2). *HvCBF7* and *TmCBF7* (Skinner et al. 2005; Miller et al. 2006) were not included in this phylogenetic analysis since their protein sequence contains a less conserved *CBF* signature and the presence of motifs found in subgroup IIIId of ERF proteins (Nakano et al. 2006).

The phylogenetic analysis presented in Fig. 1 shows that monocot *CBFs* cluster into several distinct monophyletic groups. Observation of the first group (named CBF1) reveals three distinct branches. The first branch containing the *CBF* sequences from the Arecales and Zingiberales orders (*SpCBFI*, *RhCBFI*, *DICBFI* and *ZoCBFI*) is separated from the two branches containing Poales *CBF* sequences. As additional *CBF* genes are identified and characterized in other monocotyledon plants, it may reveal functional differences that will necessitate a classification as a distinct subgroup to better reflect their evolutionary relationships. The second branch contains the lone rice gene *OsCBFI-1F*. If orthologs of this gene are found in other Poales members, this will support the need to define a separate subgroup representing these more distantly clustered genes. The remaining branch, tentatively named CBF1a, contains sequences from the Poales/Poaceae subfamilies Oryzaceae, Panicoideae and Pooideae suggesting that the ancestral Poaceae *CBFIa* was present before divergence of these subfamilies, and that none of these families have lost this *CBF* gene. In fact, Oryzaceae and Pooideae may have already possessed two genes before divergence since rice and barley each have a pair of genes (Fig. 1). Although only *TaCBFIa-A11* has been identified to date, the above information suggests that wheat could have one or two additional genes (orthologs of *CBFIa-1* and *OsCBFI-1F*). The characteristic of group CBF1 is that it is the only one that contains *CBF* genes from the three orders Poales, Arecales and Zingiberales. Although the identification of different *CBF* genes from orders other than Poales is still limited, their clustering with the CBF1 group suggests that it is the most ancient group in monocots.

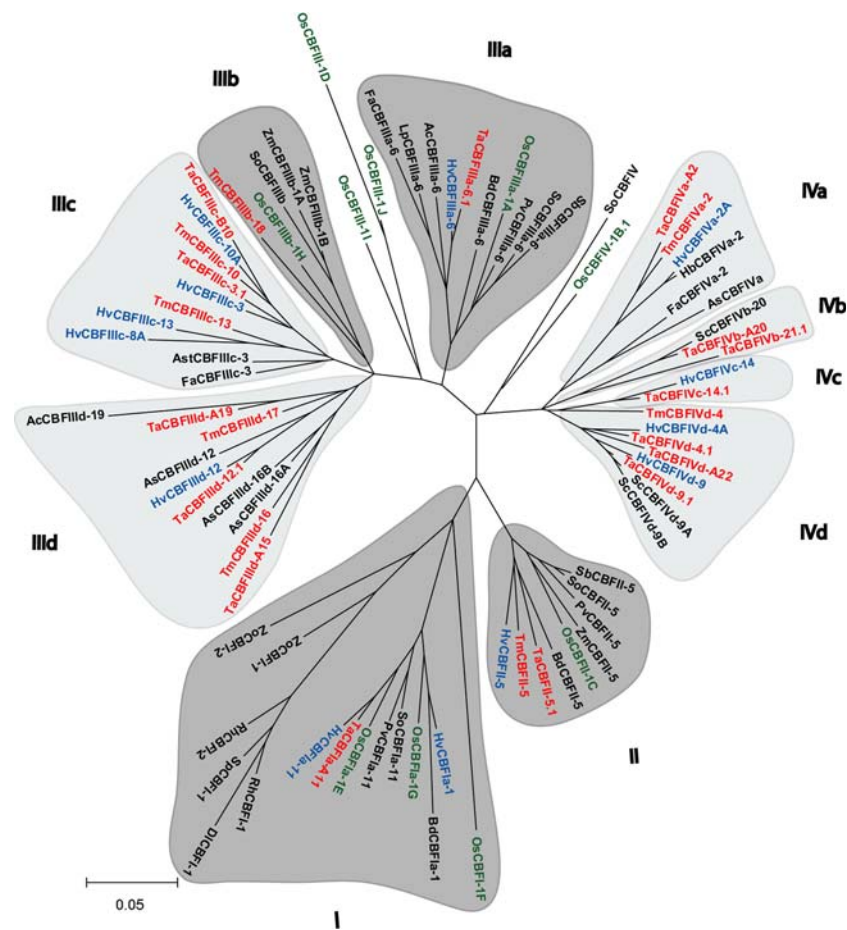


Fig. 1 Phylogenetic relationships between monocot *CBF* genes. The nucleotide sequence corresponding to the AP2 DNA binding domain and the conserved flanking signature sequences PKK/RPAGRxKFXETRHP and DSAWR defined by Jaglo et al. (2001) were aligned using ClustalW and manually adjusted. An unrooted Minimum Evolution tree was derived from this alignment using the Kimura 2-parameter. The *CBF* nomenclature used is described in Table 2. *CBF* genes belonging to specific monophyletic groups are contoured. Complete *CBF* names are shown in red for wheat species, in blue for barley, green for rice and black

for all other species examined. Groups lightly shaded contain only Pooideae sequences. *TmCBFIVd-4* was included in group IVd based on the analysis of the complete sequence. *Ac Agrostis capillaries*, *As Avena sativa*, *Ast Agrostis stolonifera*, *Bd Brachypodium distachyon*, *Dl Dypsis lutescens*, *Fa Festuca arundinacea*, *Hb Hordeum brevisubulatum*, *Hv Hordeum vulgare*, *Lp Lolium perenne*, *Os Oryza sativa*, *Pv Panicum virgatum*, *Rh Rhipidophyllum hystris*, *Sb Sorghum bicolor*, *Sc Secale cereale*, *So Saccharum officinarum*, *Sp Sabal palmetto*, *Ta Triticum aestivum*, *Tm Triticum monococcum*, *Zm Zea mays*, *Zo Zingiber officinale*

In support of this, proteins encoded by *CBFI* genes are the ones that show the highest homologies with dicotyledonous *CBF* proteins suggestive of their closer evolutionary relationship with an ancestral type *CBF*. This was also noted previously by Skinner et al. (2005).

The second group (named *CBFII*) also contains *CBF* genes from the Oryzoaceae, Panicoideae and Pooideae subfamilies but the presence of only one gene in rice suggests that it was less complex before divergence. The possibility that additional genes may exist in wheat based on the lower homology of *TmCBFII-5* (86.5%) with *T. aestivum* genes will require additional sequence identification and characterization to be confirmed. Although these genes were initially presented as part of group I (Skinner et al. 2005), they are

classified separately here as a group II based on their evolutionary distance from the first group, their specific occurrence in all Poales/Poaceae subfamilies examined, and structural differences (see next section).

The results presented in Fig. 1 show that the 11 *CBF* genes from group III are clustered in several distinct subgroups. These subgroups were named *CBFIIIa*, *CBFIIIb*, *CBFIIIc* and *CBFIIId* to reflect their monophyletic origins. Groups IIIa and IIIb are the only ones that contain *CBF* genes from the three Poaceae subfamilies suggesting that the ancestral *CBFIIIa* and *CBFIIIb* genes were already present before divergence of these subfamilies. However, group *CBFIIIb* did not always form a monophyletic clade with other substitution models suggesting a less certain orthologous

relationship. With these alternative models, *TmCBFI-Iib-18* was found to cluster alone or in the vicinity of the CBFIIIId group (results not shown). Sequencing of additional *CBFIIIb* genes will help to resolve this ambiguity. Interestingly, only *CBF* genes from tribes of the Pooideae subfamily were found clustered in groups CBFIIIc and CBFIIIId suggesting that these groups evolved following the appearance of the Pooideae. Based on the available data, group IIIc contains at least three common genes (*CBFIIIc-3*, *CBFIIIc-10* and *CBFIIIc-13*) before wheat–barley speciation. No report has yet shown the existence of the *HvCBFIIIc-8* type pseudogenes in wheat. In addition, this group contains two of the genes (*HvCBFIIIc-8* and *HvCBFIIIc-10*) that have duplicated in barley following divergence from wheat. Excluding the pseudogenes, this group is presently composed of four genes in barley and four genes in wheat. In the case of group IIIId, one barley gene has been identified (*HvCBFIIIId-12*) compared to five in wheat (*TaCBFIIIId-12*, *TaCBFIIIId-15*, *TmCBFIIIId-16*, *TmCBFIIIId-17*, *TaCBFIIIId-19*). However, it is probable that barley has at least an additional *CBFIIIId* gene since multiple homologs were identified in the more distantly related *Avena sativa*. These results suggest that differences may exist between closely related species in the exact number of *CBF* genes present within a group. The group III rice genes *OsCBFIII-1D*, *OsCBFIII-1I* and *OsCBFIII-1J* do not cluster within any of the above described subgroups. Therefore, no subgroup classification is proposed since these genes may have evolved specifically in Oryzaceae.

The analysis of the nine wheat genes from group IV (Fig. 1) reveals a compact clustering compared to group III genes, suggesting a more recent diversification. However, based on the origin of the main branches, it is possible to classify the wheat genes into four groups: CBFIVa (*TaCBFIVa-2* and *TmCBFIVa-2*), CBFIVb (*TaCBFIVb-20* and *TaCBFIVb-21*), CBFIVc (*TaCBFIVc-14*), and CBFIVd (*TaCBFIVd-4*, *TmCBFIVd-4*, *TaCBFIVd-9* and *TaCBFIVd-22*). The group CBFIVb is the only one that does not contain a barley representative at the moment. Within group CBFIVd, the identity between the complete ORF of wheat–barley orthologs (91.5% for *CBFIVd-4* and 94.2% for *CBFIVd-9*) and paralogs (90.8% for wheat–wheat and 90.3% for barley–barley) are very similar indicating that this group amplified just prior to the divergence of the wheat–barley lineage. At present, only two genes (*AsCBFIVa* and *FaCBFIVa-2*) from the closely related Pooideae tribes Aveneae and Poeae were found to cluster with the CBFIVa group indicating that at least this group had appeared prior to the radiation of these tribes. These results suggest that the

amplification of group CBFIVd and possibly the emergence of groups IVb and IVc may represent a specific characteristic of the Triticeae tribe. In addition, groups CBFIVa and CBFIVd contain the two genes (*HvCBFIVa-2* and *HvCBFIVd-4*) that have duplicated in barley following divergence from wheat. The rice (*OsCBFIV-1B.1*) and sugarcane (*SoCBFIV*) representatives were found to be distantly related to the core group IV genes suggesting that the ancestral group IV gene continued to evolve following the divergence of the corresponding plant subfamilies. A similar observation was noted previously by Skinner et al. (2005). Based on present data, barley has at least seven genes in group IV (four original plus three specifically amplified in barley) compared to nine possible genes in wheat.

In conclusion, these analyses reveal that hexaploid wheat contains at least 15 *CBF* genes, and could contain up to 23–25 *CBF* genes. The latter estimates assume hexaploid wheat has retained orthologs of all *T. monococcum* *CBF* genes and of *HvCBFIa-1* and *OsCBFI-1F*. The Poaceae *CBF* genes can be divided into ten groups with six of these (CBFIIIc, IIIId, IVa, IVb, IVc and IVd) having evolved only in the Pooideae.

Bioinformatic analysis of wheat CBF proteins

The *TaCBF* genes identified in this study encode for proteins ranging from 202 to 290 amino acids that share homologies with other CBF proteins. Analysis of these protein sequences reveals that they contain, to different degrees, the characteristic motifs found in this family (Wang et al. 2005; Skinner et al. 2005; Nakano et al. 2006). From the N- to C-terminus, we can identify the AP2 DNA-binding domain flanked by the CMIII-3 motif, and then the CMIII-1 motif, the CMIII-2 motif, and the conserved C-terminus LWSY motif (or CMIII-4). Analysis of the ORF of homeologous copies reveals that four groups have a member with a sequence difference at the C-terminus. The *TaCBFIVa-2.2* protein contains an extension of 30 amino acids past the last motif because of a T→C transition that destroys a termination codon. In the *TaCBFIIIc-3.1* gene, a transversion in the last codon creates a premature termination codon and truncates the motif to LWS. The *TaCBFIVb-A20* protein contains an extension of five amino acids comparatively to other members of this group because of a deletion of the termination codon. In the *TaCBFIVd-B22* gene, a 38 base insertion downstream of the region encoding CMIII-1 motif changes the reading frame in the remainder of the sequence thus destroying the motifs CMIII-2 and -4. How these changes impact the activity of these proteins remains to be established.

Previously, Wang et al. (2005) had noted that the occurrence of hydrophobic clusters (HCs) in the activation domain of CBF proteins were evolutionarily conserved and demonstrated their functional importance and redundant nature. To identify some of the structural differences associated with the ten CBF groups classified by phylogenetic analysis, HC analysis was performed to highlight changes to internal faces of regular secondary structures (α -helices or β -strands) that may impact their functional activity. In groups that contain only one or two members, rice and barley orthologs were included in the analysis to examine the extent of structure conservation over evolutionary time. Two regions were analyzed and include the AP2 DNA-binding domain with the motifs CMIII-3 and CMIII-1 flanking it, and the C-terminal activation

domain region containing motifs CMIII-2 and CMIII-4. For comparative purposes, we included *RhCBFI-1* and *ZoCBFI-1* from the monocotyledon orders Arecales and Zingiberales, respectively, the lone rice group I protein *OsCBFI-1F*, and *AtCBF3* from the eudicot *Arabidopsis thaliana*.

Analysis of the AP2 DNA binding region reveals that the number/positions of HCs are relatively well conserved in the different CBF protein groups (Fig. 2). These clusters correspond to previously defined regions involved in β -strand and α -helix formation in the AP2 domain (Allen et al. 1998). HC1 and HC5 show a high conservation in their length among the different groups. Some small qualitative differences are observed in the AP2 DNA binding region when comparing the ten groups of CBF proteins. For example,

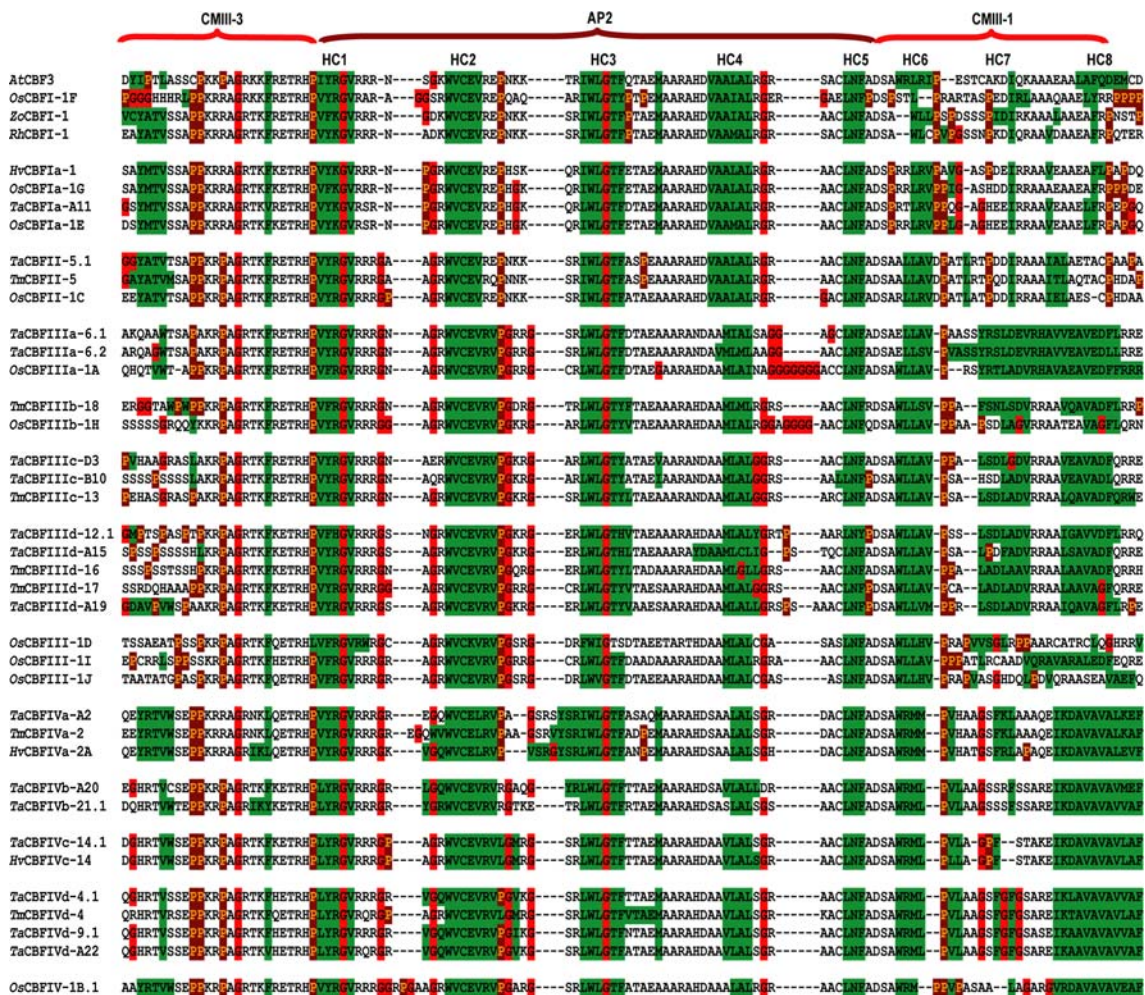


Fig. 2 Comparative hydrophobic cluster analysis of the AP2 DNA binding domain of wheat CBF protein groups. The protein sequence of the AP2 DNA binding domain and the regions CMIII-3 and CMIII-1 surrounding the AP2 domain of wheat, barley and rice CBF proteins were aligned using ClustalW and clusters of hydrophobic amino acids were highlighted in green, glycine in red

and proline in dark red. For comparative purposes, this analysis was also done for the *Arabidopsis AtCBF3*, *Oryza sativa OsCBFI-1F*, *Zingiber officinale ZoCBFI-1* and *Rhizophyllum hystrix RhCBFI-1* proteins. HC1 to HC8 identify hydrophobic clusters used to structurally define cereal CBF groups. Gaps (-) were introduced to maximize the alignment between all groups analyzed

HC2 is on average larger in CBFIV groups, decreases in size in CBFIII groups, and is smallest in CBF Ia and CBF II groups while an inverse relationship is seen for HC4. The HC2 and HC3 regions were found to be the most useful in specifically defining the four CBFIV groups. The CBFIVa group specifically contains the largest HC3 while the CBFIVd group contains the largest hydrophobic character around the HC2 region. In the case of CBFIVb, HC2 is extended more towards HC1 comparatively to the CBFIVc group where it is extended towards HC3. The presence of glycine (G) and proline (P) residues is also a characteristic that can differentiate CBF groups. The P/G pattern between HC1 and HC2 is a characteristic of CBF Ia proteins while the one between HC2 and HC3 is strictly conserved in the four CBFIII groups and the remaining group III rice proteins. The proline between HC2 and HC3 is lacking only in groups CBFIVb and CBFIVc.

Analysis of the regions surrounding the AP2 DNA binding domain reveals significant differences between different CBF groups (Fig. 2). In the CMIII-1 motif, the presence and length of HC and the P/G pattern are capable of differentiating between the CBF Ia and CBF II groups, between the CBF IIIa and the remaining CBF III groups, and between the four CBFIV groups. In the CMIII-3 region, the absence of a P is specific for the CBF Ia and CBFIVa groups, and a larger HC defines groups CBF Ia, CBF II and CBFIVa. These results show that all CBF groups besides CBF IIIb, CBF IIIc and CBF III d may have slightly different folding of their secondary and/or tertiary structures of their DNA binding domains.

Analysis of the C-terminal activation domain region (CMIII-2 and CMIII-4 motifs) reveals that CBF Ia and CBF III proteins contain fewer but larger HCs compared to CBF II and CBFIV proteins which contain a greater number of shorter HCs (Fig. 3). In the vicinity of motif CMIII-2, CBF Ia proteins contain only one HC bordered by P (Fig. 3). This structure is also found in group CBF IIIa. However, this group can be differentiated from group Ia by a different G pattern in motif CMIII-2, the absence of a HC upstream of CMIII-2, and the only CBF III group to contain a P in motif CMIII-4. CBF IIIb and CBF III d groups contain a larger HC than CBF IIIa, and share a similar structural pattern from the CMIII-2 motif to the end of the protein. However, CBF III d group can be differentiated by the presence of a HC upstream of CMIII-2. Unique features that characterize the CBF III c group include a P that creates two HCs within motif CMIII-2, and an additional HC downstream of this motif.

The CBF II and CBFIV groups contain between 4 and 6 HCs in their C-terminal activation region which

is similar to the number identified in *Arabidopsis* (Wang et al. 2005). However, the structural patterns are different. In fact, the lone protein *OsCBFI-1F* is the one that shows the highest structural similarity with *AtCBF3* suggesting that it may represent a closer version of the ancestral type CBF in monocots. The CBF II group contains five HCs and is unique among the groups analyzed since it does not contain P at the end of motif CMIII-2. The CBFIVa and CBFIVd groups share a similar structure with three HCs in motif CMIII-2 and up to five common HCs in total. Group CBFIVd can be differentiated by the presence of the first HC upstream of motif CMIII-2 in all members and a P in motif CMIII-4. Groups CBFIVb and CBFIVd contain only two HCs in motif CMIII-2, and these display variable lengths. In addition, group CBFIVc contains an additional HC downstream of CMIII-2 which is not found in any other group IV CBF. In summary, these results show that most groups described in this study display some structural differences in the AP2 DNA binding region while all groups show differences in their C-terminal activation regions. An important observation from these results is that rice proteins (*OsCBFIII-1D*, *OsCBFIII-1I*, *OsCBFIII-1J* and *OsCBFIV-1B.1*) do not show clear structure conservation with any of the described groups corroborating their non-orthologous relationship. This is in contrast to the rice members of groups CBF Ia, CBF II and CBF IIIa and partly *ZoCBFI-1* and *RhCBFI-1* that do show structures that are orthologous in nature, and have been conserved since the divergence of the respective branches.

Expression of wheat *CBF* genes

CBF genes are known to be induced rapidly upon exposure to LT. To determine the expression behaviour of the 15 *TaCBF* gene groups identified in this study, we initiated a LT time course using two cultivars differing in their FT capacities. The LT treatment was initiated 4 h after dawn and continued for 7 days. Probes outside of the AP2 DNA binding domain were designed for northern blot analyses to avoid cross reaction with known *TaCBF* genes (Table S3). The results obtained with *TaCBFIVb-21.1* and the genes from groups *CBFIa* and *CBFIIIc* are not included in Fig. 4 since no signals above background were detected. This suggests that these genes are expressed at low levels under the conditions assayed. The remaining 11 genes displayed signals and are shown in Fig. 4. Analysis of these results allows two general observations to be drawn. The first is that all assessed and detectable *TaCBF* genes display a transitory induction profile. In

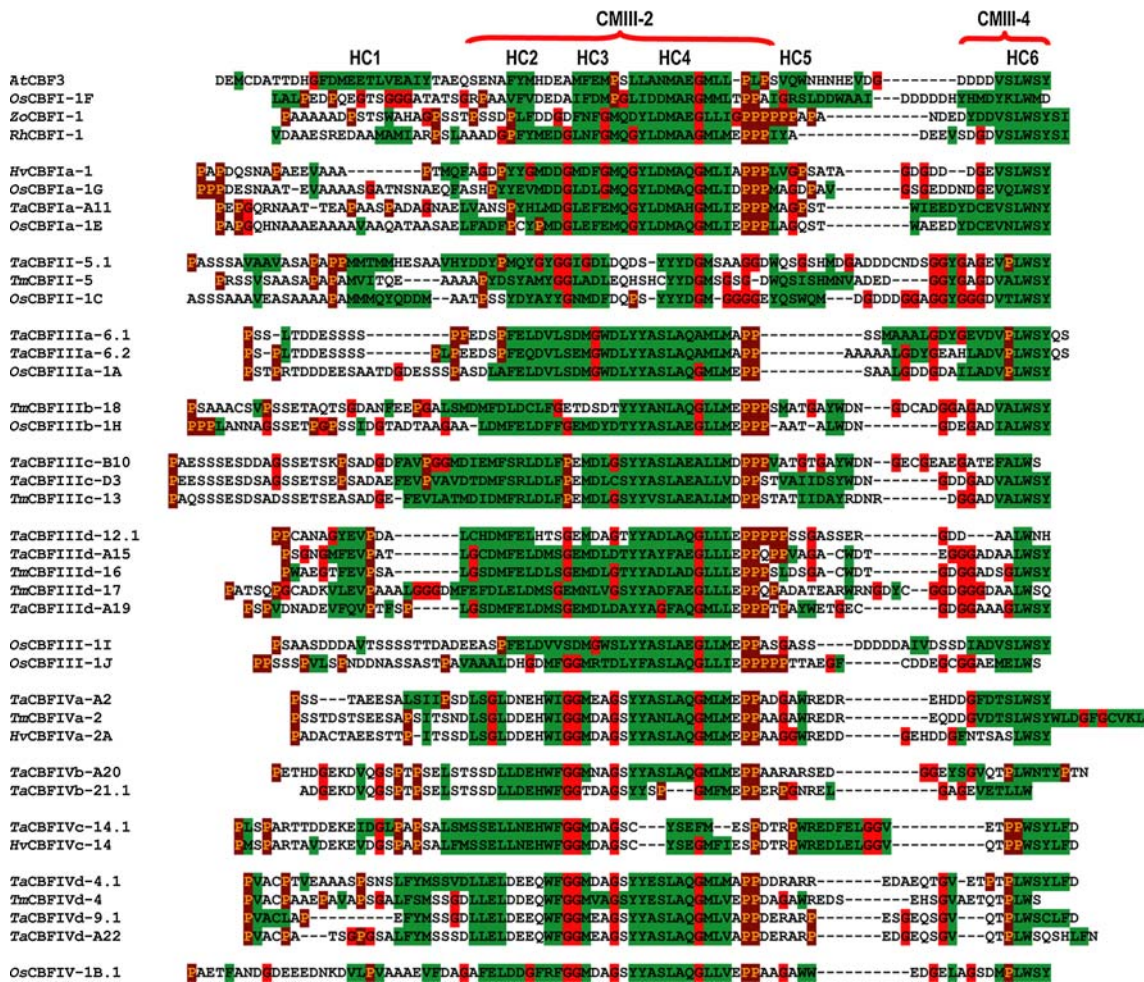


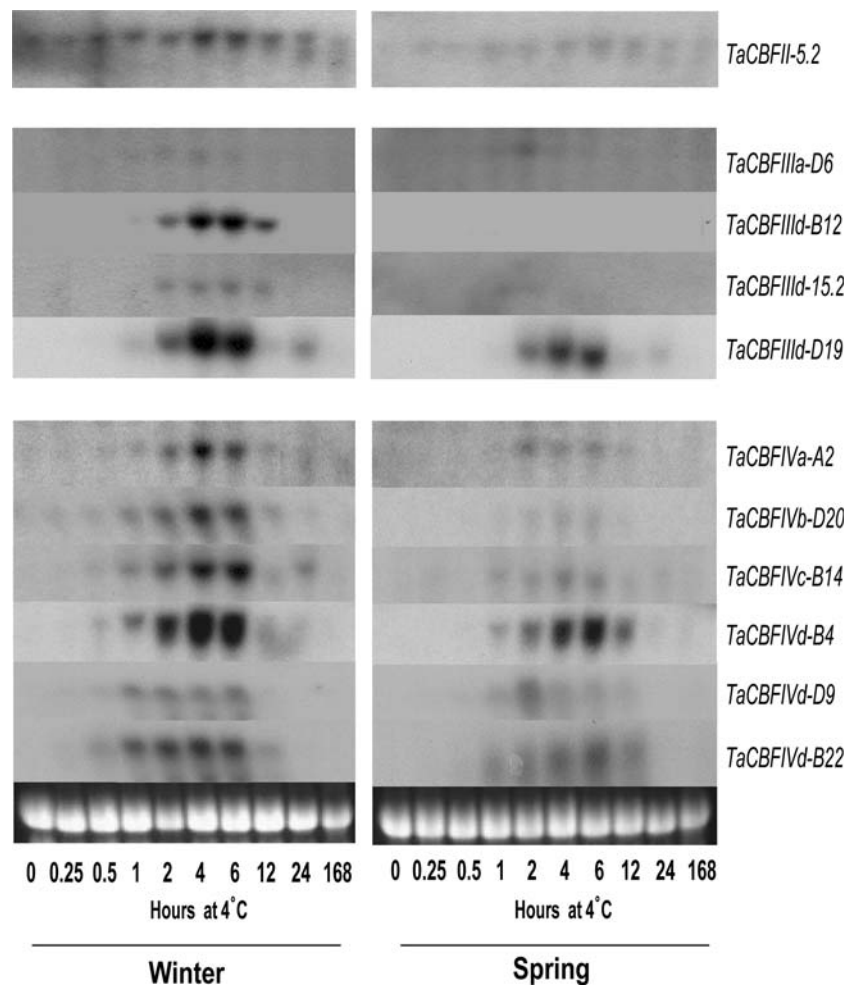
Fig. 3 Comparative hydrophobic cluster analysis of the C-terminal region of wheat CBF proteins. The regions surrounding motifs CMIII-2 and CMIII-4 of wheat CBF proteins were aligned using ClustalW and clusters of hydrophobic amino acids were highlighted in green, and glycine in red and proline in dark red. For comparative purposes, this analysis was also done for the *Arabidopsis*

AtCBF3, *Oryza sativa* *OsCBFI-1F*, *Zingiber officinale* *ZoCBFI-1* and *Rhipidophyllum hystrix* *RhCBFI-1* proteins. HC1 to HC6 identify hydrophobic clusters previously defined in *Arabidopsis* (Wang et al. 2005). Gaps (-) were introduced to maximize the alignment between members of a group, and to maximize the comparison of similar regions between groups

fact, all *TaCBF* genes showed little or no expression before the onset of the LT treatment. They were induced by LT and attained maximum levels after 4–6 h of treatment, and then returned to basal levels after 1–7 days of treatment. The second observation revealed by our northern analyses is that 9 of the 11 *TaCBF* genes assayed are expressed to higher levels in the winter cultivar compared to the spring cultivar suggesting that higher *TaCBF* expression is associated with the winter cultivar's superior FT development capacity. However, there are differences in the quantitative accumulation of certain *TaCBFs*. For example, the expression of *TaCBFIIId-B12* and *TaCBFIIId-15.2* was not detected in the spring cultivar while the remaining seven genes were expressed at low levels compared to the winter cultivar.

Since the northern expression study on *TaCBFs* (Fig. 4) was not sensitive enough to measure the basal or inheritable expression of *TaCBFs*, we decided to use the more sensitive quantitative real-time PCR to quantify the initial LT response in a winter and spring cultivar. The experiment was designed to include several consecutive 20 and 4°C time points for evaluating the extent of inheritable versus LT inducible fluctuation, and to initiate the LT treatment near the end of the day to better reflect natural conditions. This experimental design also allowed a comparison of the influence of day period on *TaCBF* induction by LT. Primer sets used in these experiments were designed against only one copy of the 15 *TaCBF* gene groups identified (Table S4) and shown to be specific by mapping these genes to unique chromosome arms (Table 1). To

Fig. 4 Northern analyses of 11 *TaCBF* transcripts during cold treatments of winter and spring wheats. Total RNA was extracted from wheat leaves and analyses performed as described in the [Materials and methods](#). Hybridizations were done on a series of replicate blots with unique probes designed outside of the AP2 DNA binding domain of 15 different *CBF* genes (Table S3), and the 11 *TaCBF* transcripts that displayed signals are shown. The cold treatment (4°C) was initiated 4 h into the day phase and continued for the times indicated. The ethidium bromide stained rRNA load that is shown is representative of each gel transferred to a membrane. The winter and spring cultivars were Norstar and Quantum, respectively



compare the *CBF* induction patterns in a winter and spring cultivar, panels in Fig. 5 were generated using the independent calibrators winter 08:00 and spring 08:00, respectively. This resulted in y-axis scales that are not directly comparable among cultivars but allow the direct visualization of the LT effect on gene expression. Once these profiles were obtained, the relative quantity of winter versus spring LT accumulation was experimentally determined after 2 h of LT exposure at 22:00 (shown as the winter/spring expression at 22:00 in Fig. 5) while the basal expression value was measured for the control point 18:00 (results not shown). From the expression pattern of *TaCBF* genes, 13 showed statistically significant basal and/or LT expression (Fig. 5). The exceptions that did not show any reproducible signal were *TaCBFIIa-A11* and *TaCBFIIIc-B10* in both cultivars (results not shown).

Results of Fig. 5 show that in almost all cases the maximum accumulation of *TaCBFs* occurs after 2 h of LT treatment. These results contrast with those of 4 and 6 h observed in Fig. 4 and suggest that the pattern of induction is influenced by the period of the day

when the LT treatment was initiated. Several other observations can be noted from the results presented in Fig. 5. *TaCBFIIIc-D3* displays an extreme transient expression profile showing low expression in all points examined except the 2 h LT treatment. A similar expression profile was observed for the barley ortholog (Choi et al. 2002) suggesting evolutionary conservation of regulation pattern. The fact that no detectable expression of *TaCBFIIa-A11* and *TaCBFIIIc-B10* was observed in this study suggests that *CBFIIa* and *CBFIIIc* genes may be expressed under specific conditions and/or extremely low levels. Other observations include: a near constitutive expression for *TaCBFII-5.2*, a low and transient induction for *CBFIIIa* and *CBFIVb* genes, and a high and more sustained expression for most *CBFIIIc*, *CBFIVc* and *CBFIVd* genes. A comparison of winter and spring LT induction profiles reveals that they are qualitatively very similar. The quantitative comparison of the 2 h LT time points reveals that *CBFII*, *CBFIIIa* and *CBFIIIc* genes are expressed to similar levels (within a threefold factor) in both cultivars. On the other hand, *CBFIIIc*, *CBFIVa*, *CBFIVb*,

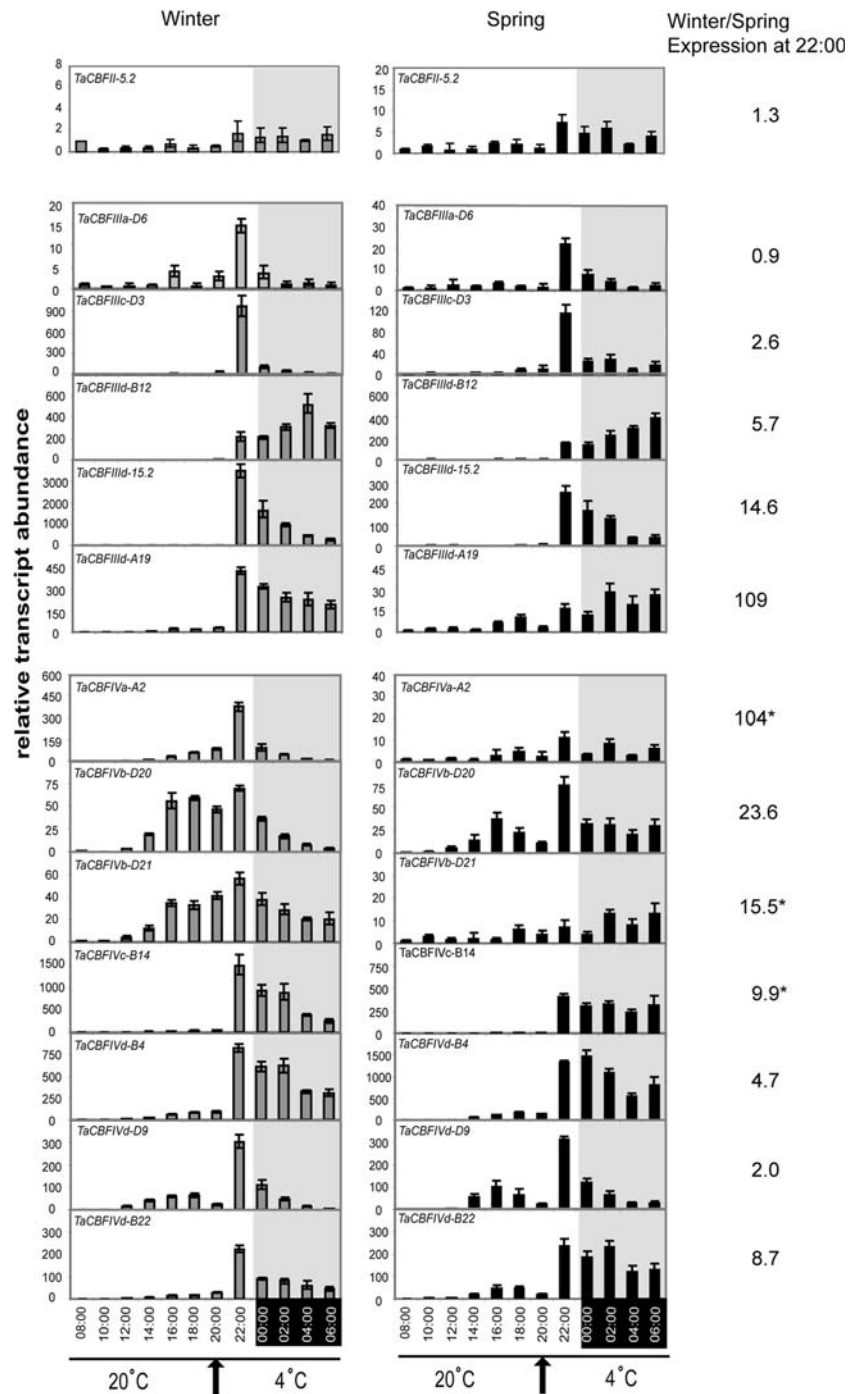


Fig. 5 Quantitative real-time PCR expression analysis of *TaCBF* genes in two wheat cultivars in response to low temperature. Plants were germinated for 7 days at 20°C under a 16-h-day/8-h-night photoperiod. Beginning on day 8 at 8:00, plants were grown at 20°C for 12 h (from 08:00 to 20:00) then exposed at 4°C for 10 h (20:00–06:00). *Gray areas* represent the dark periods and *arrows* the start of the LT treatment. Leaf samples were harvested at each indicated time point and RNA extracted and analyzed by quantitative real-time PCR as described in the [Materials and methods](#). Relative transcript abundance was calculated and normalized with respect to the 18S rRNA transcript level. Winter and spring panels were generated using the independent calibrators winter 08:00 and spring 08:00, respectively. This resulted in

y-axis scales that are not directly comparable among cultivars but allow the direct visualization of the LT effect on gene expression within each cultivar. Error bars indicate the range of possible RQ values defined by the SE of the delta threshold cycles (Cts). The winter and spring cultivars were Norstar and Manitou, respectively. Winter/Spring Expression at 22:00 was determined by directly comparing the RQ values in both cultivars at the 22:00 time point. The RQ_{max} and RQ_{min} values did not exceed 25% deviation of the value shown except as indicated by (*asterisks*) where the deviation ranged between 36 and 45%. It is noted that a different genome equivalent was assayed in this experiment (*TaCBFIIId-A19*) compared to Fig. 4 (*TaCBFIIId-D19*) which prevents direct comparisons

CBFIVc and *CBFIVd* genes (except *TaCBFIVd-D9*) show increased LT expression (4.7-fold and more) in the winter compared to the spring cultivar. In addition, these five groups (except for *TaCBFIIIId-B12*) also show a higher basal expression in the winter cultivar (results not shown), and this can be easily evaluated for some members when comparing the induction profiles in Fig. 5 if we consider that the LT expression of these genes is more than 4.7-fold higher in winter compared to spring cultivars as indicated above. These results indicate that the five CBF groups are associated with both the superior inherited and LT inducible capacities of the winter cultivar to develop FT.

An additional interesting observation that emerged from this experiment is that the expression of several *TaCBF* genes was not constant during the seven 20°C time points indicating that cold treatment is not needed for their expression. The expression of these genes was high in the vicinity of 8–10 h after dawn and then decreased in both winter and spring cultivars (for example see *TaCBFIVb-D20* in Fig. 5). To better visualize this behaviour, the 20°C experiment was repeated with the cultivar Norstar for a full 24 h period without cold treatment (Fig. 6). These results reveal that genes from the four CBFIV groups and two of the three genes from the CBFIIIId group show a diurnal fluctuation in their expression with maxima appearing between 8 and 14 h after dawn and minima between 20 and 24 h after dawn under long day conditions. This diurnal fluctuation is reproducible since it was observed for two additional cycles in experiments with *TaCBFIVb-D20*, *TaCBFIVc-B14*, *TaCBFIVd-B4*, *TaCBFIVd-D9* and *TaCBFIVd-B22* (results not shown). Therefore, the five groups that were found to be more expressed in the winter cultivar also display a characteristic diurnal fluctuation during growth at 20°C.

Discussion

Wheat is a good model species for studying FT since its tolerance lies between that of freezing sensitive plants (rice, maize and oat) and the extremely tolerant species rye. To decipher the genetic basis underlying the different capacities of temperate cereal species in developing FT, we have initiated the identification of genes that have the potential to influence FT. Since *CBF* genes have been widely implicated in cold acclimation in many species, we identified and characterized 15 *TaCBF* gene groups in hexaploid wheat. Our analyses reveal that wheat species, *T. aestivum* and *T. monococcum*, may have a large and complex *CBF*

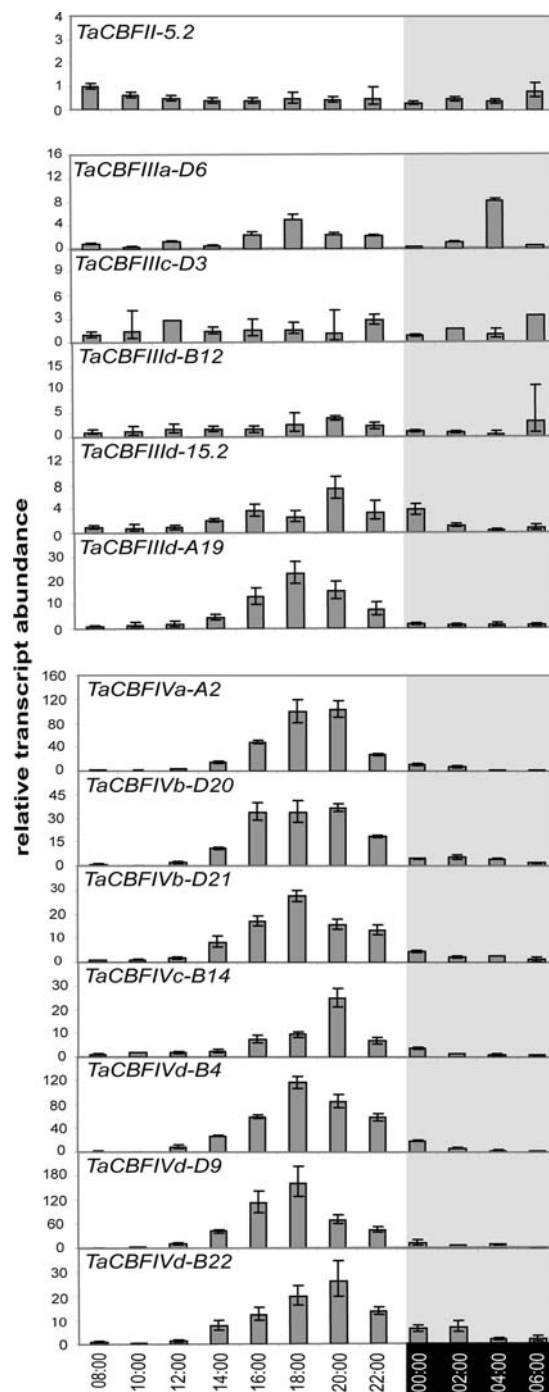


Fig. 6 Quantitative real-time PCR expression analysis of *TaCBF* genes in Norstar in response to a diurnal cycle. Plants were germinated for 7 days at 20°C under a 16-h-day/8-h-night photoperiod. Beginning on day 8 at 08:00, plants were grown at 20°C for 22 h (from 08:00 to 06:00). Gray areas represent the dark periods. Leaf samples were harvested at each time point, and RNA was extracted and analyzed by quantitative real-time PCR as described in the [Materials and methods](#). Relative transcript abundance was calculated and normalized with respect to the 18S rRNA transcript level and the calibrator time point (08:00). Error bars indicate the range of possible RQ values defined by the SE of the delta threshold cycles (Cts)

family with up to 25 different *CBF* genes. The large number of *CBF* genes in wheat is comparable to the number found in barley (20 or more) (Skinner et al. 2005) but contrasts with the ten genes present in rice and six in *Arabidopsis*. It is not known why freezing tolerant cereals have evolved and maintained so many *CBF* genes. Since the amplification of the *CBF* gene family has evolved independently after the monocot–eudicot divergence (Qin et al. 2004; Bräutigam et al. 2005; Xiong and Fei 2006), a thorough characterization of a large number of *CBF* genes in wheat seemed a daunting task. At present, it is difficult to assess if all *TaCBFs* have specific functions, obtained by subfunctionalization and neofunctionalization, or if they all have redundant functions. Therefore, an important objective of this study was to classify *CBF* genes into functional categories that would help orient functional studies. Towards this goal, we studied the evolution of the *CBF* family in monocotyledons and determined their structural characteristics using HCA to display conservation/changes that could affect protein secondary structure.

This study indicates that the *CBF* amplification seen in wheat has occurred quite recently and following the emergence of the Oryzaceae, Panicoideae and Pooideae lineages. From the phylogenetic analysis, it is easily observable that these subfamilies had already evolved representatives of groups Ia, II, IIIa and possibly IIIb suggesting that orthologs within these groups should have common functions. The HCA analyses confirm that rice and wheat orthologs have similar structural characteristics that have been conserved over evolution. On the other hand, groups IIIc, IIIId, IVa, IVb, IVc and IVd representing 18 genes in wheat species arose following the emergence of the Pooideae since no rice or maize *CBFs* clustered within these groups. After the emergence of Oryzaceae, rice only gained four genes. However, these rice genes are distantly related to the groups containing wheat genes and their proteins do not display similar structural features. Therefore, the 18 wheat genes in groups IIIc, IIIId, IVa, IVb, IVc and IVd and the 4 unclassified group III and IV rice genes may represent the *CBF* response machinery that evolved in Pooideae and Oryzaceae, respectively, as they radiated into specific habitats. Since these arose after the subfamilies split, it is not surprising to see that structural patterns are not conserved in the respective *CBF* proteins. There is some indications that the *CBF* family is still (or has been recently) evolving under some selective pressure. The first comes from the observation that, although the evolutionary distance between groups IVa, IVb, IVc and IVd is small, their proteins display notable structural differences in their AP2 domain and C-terminal region. This

is evident between groups but can also be seen between members of group IVa (Fig. 2) and IVb (Fig. 3). Finally, the observation that barley contains several duplicated genes with high similarities (Skinner et al. 2005) suggests that they arose after the wheat–barley divergence with some (those with identities above 95%) having arisen in the last 4 MY. As more *CBF* sequence information becomes available, it will allow a detailed evaluation of the number and nature of *CBF* gene groups found in different subfamilies of the Poaceae and even in other monocot orders. In a general sense, this will lead to a better understanding of the evolution of *CBF*-mediated tolerance to abiotic stresses, and in a more practical way, it may allow associating the emergence (or loss) of certain *CBF* genes (or groups) with maximum species freezing tolerance.

An interesting conclusion that can be drawn from the phylogenetic study is that the evolution of the Pooideae in a specific ecosystem has impacted the *CBF* signaling machinery by increasing the total number of *CBF* genes and the number of functional categories as also corroborated by the HCA analyses. This complexity is found in the Triticeae tribe and suggests that these plants have faced a strong selection pressure to maintain genes that will help them to perform well under a variety of environmental conditions. The selection pressure may not be constant but could intensify during generations that experience an unusually severe winter.

Studies with monocot *CBFs* have shown that they share the conserved domains present in this protein family, and that they are capable of binding a DRE-related *cis* element and inducing *COR* gene expression in *Arabidopsis* although with a lesser efficiency or an incomplete response (Dubouzet et al. 2003; Qin et al. 2004; Skinner et al. 2005) compared to overexpression of endogenous *Arabidopsis CBFs* (Jaglo-Ottosen et al. 1998; Liu et al. 1998). This can be partly explained by structural differences which make monocot *CBFs* less efficient in replacing the endogenous *Arabidopsis* protein function. The independent evolution of *CBF* genes in plants will certainly make it harder to elucidate the exact roles of cereal genes in model species like *Arabidopsis* that are more evolutionary distant. Therefore, to understand the exact functions of Pooideae- and possibly Triticeae-specific groups/genes, it will be essential to study these *CBF* genes in species such as wheat and barley. In addition, determining the exact contributions of members of the *CBF* family may be even more complex than anticipated since members of other subgroups of group III ERF proteins have been shown to be cold regulated and capable of binding a DRE *cis* element such as *HvCBF7* from barley

(Skinner et al. 2005) and TINY2 from *Arabidopsis* (Wei et al. 2005). Their contribution to the regulation of *COR* gene expression in species with large *CBF* families or their possible compensation in species with small families needs to be explored.

The HCA analysis of the AP2 DNA binding domain was capable of differentiating seven out ten groups suggesting that protein structure and binding properties may be affected. Results of Xue (2002) demonstrated that *HvCBFIa-1* preferred TTGCCGACAT as a binding site while *HvCBFIVa-2* (Xue 2003) preferred YYGTCGACAT. In addition, *HvCBFIVa-2* and *HvCBFIVd-4A* showed a LT dependence for maximal binding activity (Xue 2003; Skinner et al. 2005). These results corroborate that functional differences can be visualized through HCA analyses and allow us to predict that up to seven groups may show some differences in DNA binding properties while group members will show a redundancy in binding. Determining such properties will be important for understanding the possible differences/overlap in regulons controlled by *CBF* groups. It was recently suggested that small variation in transcription factor binding consensus could have important consequences for bioactivity (Benedict et al. 2006) and some of the barley *CBFs* have been demonstrated to have differential affinity for specific CRT/DRE motifs (Xue 2002, 2003; Skinner et al. 2005). On the other hand, the HCA analysis of the C-terminal activation domain revealed substantial differences between all ten groups. The structural patterns were relatively well conserved between group members even for those corresponding to distantly related species. The varied patterns detected may be at the base of specific functional differences which could impact protein folding, recognition of specific interaction partners and the transactivation potential of a specific set of *CBF* proteins. The molecular dissection of *Arabidopsis* *CBF1* (Wang et al. 2005) showed that certain hydrophobic clusters and other structural determinants in this region were important in regulating transactivation potential of this region. In *Brassica napus*, two closely related *CBF* proteins were shown to have substantially different transactivation potentials in yeast and tobacco (Zhao et al. 2006). The major differences between these proteins lie in the C-terminal activation domain within the hydrophobic clusters. Such behaviour has not been reported for the *CBFI*, 2, and 3 genes of *Arabidopsis* (Gilmour et al. 2004) which diverged from *Brassica* some 24 MYA (Koch et al. 2000) suggesting that even related species may have evolved some specific differences in *CBF* properties. These examples illustrate that the C-terminal activation domain of different groups could have distinct properties that play

important and specific roles. In addition, this last observation suggests that some *CBF* properties in group IVa (Fig. 2) and IVb (Fig. 3) may even differ between the closely related species barley and wheat. The three structurally different groups IIIc, IIId and IVd share the characteristic of having several members with similar structural patterns which contrast with the lower complexity present in other groups. An explanation for the selection and conservation of duplicated genes that seem to accomplish redundant functions comes from work in yeast and suggests a selection for a higher flux through the pathways controlled by these genes (Papp et al. 2004). Therefore, the wheat genes from these groups may be necessary in certain circumstances where maximal induction is needed to achieve very high levels of LT tolerance and/or activation of a large number of genes in a regulon.

Quantitative RT-PCR analyses revealed that the expression level of groups IIIId, IVa, IVb, IVc and IVd was more pronounced in the winter cultivar (fourfold and more) compared to the remaining groups assayed (threefold and less). This was also demonstrated in barley and wheat for members of the *CBFIV* groups (Kume et al. 2005; Skinner et al. 2005). It is probably not a coincidence that five of the six groups that evolved in the Pooideae show expression levels that are correlated with the winter cultivar's capacity to develop LT tolerance. Previous studies had already noted that the majority of the expansion of *CBF* genes has occurred from an ancestral cluster/locus (Skinner et al. 2005, 2006; Miller et al. 2006). In rice, three *CBF* genes (*OsCBFIIIa-1A*, *OsCBFIIIb-1H* and *OsCBFIV-1B.1*) are present as a tandem cluster on a region on rice chromosome 9 that is collinear with the chromosome 5 region of the Triticeae where the *CBFs* occur. Therefore in the Triticeae (*T. monococcum*, barley and hexaploid wheat), there has been amplification of the genes in this region to give rise to the six groups present specifically in this tribe. These observations suggest that as FT-associated *CBF* genes were amplified, selective pressure has maintained their role in FT. The genetic capacity of the winter cultivar to induce a higher level of expression during the LT response is also reflected in constitutive levels measured under control growth conditions. Since LT is not involved in this regulation, the higher level in winter versus spring wheat comes from different inheritable capacities to express *CBF* genes. A similar association was observed between the inheritable level of *CBF* expression and the LT tolerance of different *Arabidopsis* lines collected at different latitudes (Hannah et al. 2006). These observations are not surprising since a study had already shown an association between higher

levels of the WCS120 protein family and cultivar capacity to develop FT (Houde et al. 1992).

The expression patterns also revealed that groups IIIId, IVa, IVb, IVc and IVd displayed a diurnal fluctuation that peaked 8–14 h after dawn. This natural rhythm influenced the time course of LT induction with faster inductions during evenings (maximum at 2 h) and slower induction during mornings (maximum at 4–6 h). Although the peak period of induction does not coincide with the coolest period of the day, it does coincide with the daily decrease of temperature during sunset suggesting the rhythm is preceding/anticipating the event. Circadian clock regulation of *CBF* expression has been described in *Arabidopsis* (Fowler et al. 2005), and therefore, may be common in plants capable of developing FT since it would confer a selective advantage during sudden drops in LT. In temperate cereals, group IV proteins (*HvCBFIVa-2* and *HvCBFIVd-4A*) were shown to bind *cis* elements in a LT-dependent manner. If this property is present in all group IV members of temperate cereals, the daily accumulation of these proteins during normal growth conditions would not cause profound changes in *COR* gene expression and thus prevent wasting cellular resources since their DNA binding activity is relatively low at this temperature. Once exposed to a sudden drop in LT, the DNA binding activity of these factors would increase, and this would immediately impact *COR* gene expression. The accumulation of partially inactive factors during warm growth conditions would also alleviate deleterious symptoms from developing as observed in transgenic plants constitutively overexpressing complete or portions of *CBF* genes (Liu et al. 1998; Wang et al. 2005; Ito et al. 2006). Therefore, group IV proteins from temperate cereals may ultimately represent a uniquely engineered protein group that functions as a first line of defence against sudden drops in LT. The functional studies of group IV CBFs is thus essential to understand the specific contribution of these groups and their impact on the range of LT tolerance capacities observed in temperate cereals. In addition, these studies may identify unique properties that could be incorporated in CBF proteins from other species.

In conclusion, this study reveals that wheat species, *T. aestivum* and *T. monococcum*, may contain up to 25 *CBFs*. These genes can be divided into at least ten groups that share a common phylogenetic origin and similar structural characteristics. Six of these groups (CBFIIIc, IIIId, IVa, IVb, IVc and IVd) are found only in the Pooideae, suggesting that they evolved recently during the colonization of temperate habitats. Expression studies revealed that five groups (CBFIIIId, IVa, IVb,

IVc and IVd) display higher constitutive and LT inducible expressions in the winter cultivar. The higher inherited and inducible *CBF* expression suggests that these groups may be major components that regulate the capacities of Pooideae species to develop LT tolerance.

Acknowledgments We wish to thank F. Allard, M. Champoux, N. Toudji and K. Tremblay for their technical assistance. The excellent editing work of one of the anonymous reviewers is also appreciated. This study was supported by Genome Canada and Génome Québec grants (F.S. and M.H.), and by Natural Sciences and Engineering Research Council of Canada discovery grant (F.S. and J.D.). M.B. was supported by a scholarship from the Ministry of State for Higher Education and Scientific Research, Egypt.

References

- Allen MD, Yamasaki K, Ohme-Takagi M, Tateno M, Suzuki M (1998) A novel mode of DNA recognition by a β -sheet revealed by the solution structure of the GCC-box binding domain in complex with DNA. *EMBO J* 17:5484–5496
- Båga M, Chodaparambil SV, Limin AE, Pecar M, Fowler DB, Chibbar RN (2007) Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. *Funct Integr Genomics* 7:53–68
- Benedict C, Geisler M, Trygg J, Huner N, Hurry V (2006) Consensus by democracy. Using meta-analyses of microarray and genomic data to model the cold acclimation signaling pathway in *Arabidopsis*. *Plant Physiol* 141:1219–1232
- Bräutigam M, Lindlöf A, Zakhrebekova S, Gharti-Chhetri G, Olsson B, Olsson O (2005) Generation and analysis of 9792 EST sequences from cold acclimated oat, *Avena sativa*. *BMC Plant Biol* 5:18
- Callebaut I, Labesse G, Durand P, Poupon A, Canard L, Chomilier J, Henrissat B, Mornon JP (1997) Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. *Cell Mol Life Sci* 53:621–645
- Choi DW, Rodriguez EM, Close TJ (2002) Barley *CBF3* gene identification, expression pattern, and map location. *Plant Physiol* 129:1781–1787
- Danyluk J, Sarhan F (1990) Differential mRNA transcription during the induction of freezing tolerance in spring and winter wheat. *Plant Cell Physiol* 31:609–619
- Dubouzet JG, Sakuma Y, Ito Y, Kasuga M, Dubouzet EG, Miura S, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Os-DREB genes in rice, *Oryza sativa* L., encode transcription activators that function in drought-, high-salt- and cold-responsive gene expression. *Plant J* 33:751–763
- Endo TR (1988) Induction of chromosomal structural changes by a chromosome of *Aegilops cylindrica* L in common wheat. *J Hered* 79:366–370
- Endo TR, Gill BS (1996) The deletion stocks of common wheat. *J Hered* 87:295–307
- Fowler SG, Cook D, Thomashow MF (2005) Low temperature induction of *Arabidopsis CBF1*, 2, and 3 is gated by the circadian clock. *Plant Physiol* 137:961–968
- Francia E, Rizza F, Cattivelli L, Stanca AM, Galiba G, Tóth B, Hayes PM, Skinner JS, Pecchioni N (2004) Two loci on chromosome 5H determine low-temperature tolerance in a ‘Nure’ (winter) \times ‘Tremois’ (spring) barley map. *Theor Appl Genet* 108:670–680

- Gaboriaud C, Bissery V, Benchetrit T, Mornon JP (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett* 224:149–155
- Gill BS, Appels R, Botha-Oberholster A-M, Buell CR, Bennetzen JL, Chalhoub B, Chumley F, Dvořák J, Iwanaga M, Keller B, Li W, McCombie WR, Ogihara Y, Quetier F, Sasaki T (2004) A workshop report on wheat genome sequencing: International Genome Research on Wheat Consortium. *Genetics* 168:1087–1096
- Gilmour SJ, Fowler SG, Thomashow MF (2004) *Arabidopsis* transcriptional activators CBF1, CBF2, and CBF3 have matching functional activities. *Plant Mol Biol* 54:767–781
- Haake V, Cook D, Riechmann JL, Pineda O, Thomashow MF, Zhang JZ (2002) Transcription factor CBF4 is a regulator of drought adaptation in *Arabidopsis*. *Plant Physiol* 130:639–648
- Hannah MA, Wiese D, Freund S, Fiehn O, Heyer AG, Hincha DK (2006) Natural genetic variation of freezing tolerance in *Arabidopsis*. *Plant Physiol* 142:98–112
- Houde M, Dhindsa RS, Sarhan F (1992) A molecular marker to select for freezing tolerance in Gramineae. *Mol Gen Genet* 234:43–48
- Houde M, Belcaid M, Ouellet F, Danyluk J, Monroy A, Dryanova A, Gulick P, Bergeron A, Laroche A, Links M, McCarthy L, Crosby WL, Sarhan F (2006) Wheat EST resources for functional genomics of abiotic stress. *BMC Genomics* 7:149
- Huang X, Madan A (1999) CAP3: a DNA sequence assembly program. *Genome Res* 9:868–877
- Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002a) Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. *Plant Mol Biol* 48:805–820
- Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002b) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. *Proc Natl Acad Sci USA* 99:8133–8138
- Ito Y, Katsura K, Maruyama K, Taji T, Kobayashi M, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2006) Functional analysis of rice DREB1/CBF-type transcription factors involved in cold-responsive gene expression in transgenic rice. *Plant Cell Physiol* 47:141–153
- Jaglo-Ottosen KR, Gilmour SJ, Zarka DG, Schabenberger O, Thomashow MF (1998) *Arabidopsis* CBF1 overexpression induces *COR* genes and enhances freezing tolerance. *Science* 280:104–106
- Jaglo KR, Kleff S, Amundsen KL, Zhang X, Haake V, Zhang JZ, Deits T, Thomashow MF (2001) Components of the *Arabidopsis* C-repeat/dehydration-responsive element binding factor cold-response pathway are conserved in *Brassica napus* and other plant species. *Plant Physiol* 127:910–917
- Kellogg EA (2001) Evolutionary history of the grasses. *Plant Physiol* 125:1198–1205
- Koch MA, Haubold B, Mitchell-Olds T (2000) Comparative evolutionary analysis of chalcone synthase and alcohol dehydrogenase loci in *Arabidopsis*, *Arabis*, and related genera (Brassicaceae). *Mol Biol Evol* 17:1483–1498
- Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 5:150–163
- Kume S, Kobayashi F, Ishibashi M, Ohno R, Nakamura C, Takumi S. (2005) Differential and coordinated expression of *Cbf* and *Cor/Lea* genes during long-term cold acclimation in two wheat cultivars showing distinct levels of freezing tolerance. *Genes Genet Syst* 80:185–197
- Limin AE, Danyluk J, Chauvin LP, Fowler DB, Sarhan F (1997) Chromosome mapping of low-temperature induced *WCS120* family genes and regulation of cold-tolerance expression in wheat. *Mol Gen Genet* 253:720–727
- Liu Q, Kasuga M, Sakuma Y, Abe H, Miura S, Yamaguchi-Shinozaki K, Shinozaki K (1998) Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain separate two cellular signal transduction pathways in drought- and low-temperature-responsive gene expression, respectively, in *Arabidopsis*. *Plant Cell* 10:1391–1406
- Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta Ct}$ method. *Methods* 25:402–408
- Magome H, Yamaguchi S, Hanada A, Kamiya Y, Oda K (2004) *dwarf and delayed-flowering 1*, a novel *Arabidopsis* mutant deficient in gibberellin biosynthesis because of overexpression of a putative AP2 transcription factor. *Plant J* 37:720–729
- Miller AK, Galiba G, Dubcovsky J (2006) A cluster of 11 *CBF* transcription factors is located at the frost tolerance locus *Fr-A^{m2}* in *Triticum monococcum*. *Mol Genet Genomics* 275:193–203
- Nakano T, Suzuki K, Fujimura T, Shinshi H (2006) Genome-wide analysis of the ERF gene family in *Arabidopsis* and rice. *Plant Physiol* 140:411–432
- Novillo F, Alonso JM, Ecker JR, Salinas J (2004) CBF2/DREB1C is a negative regulator of *CBF1/DREB1B* and *CBF3/DREB1A* expression and plays a central role in stress tolerance in *Arabidopsis*. *Proc Natl Acad Sci USA* 101:3985–3990
- Papp B, Pál C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* 429:661–664
- Qin F, Sakuma Y, Li J, Liu Q, Li Y-Q, Shinozaki K, Yamaguchi-Shinozaki K (2004) Cloning and functional analysis of a novel DREB1/CBF transcription factor involved in cold-responsive gene expression in *Zea mays* L. *Plant Cell Physiol* 45:1042–1052
- Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics methods and protocols: methods in molecular biology*. Humana Press, Totowa, pp 365–386
- Sakai A, Larcher W (1987) Frost survival of plants: responses and adaptation to freezing stress. Springer, Berlin
- Skinner JS, von Zitzewitz J, Szucs P, Marquez-Cedillo L, Filichkin T, Amundsen K, Stockinger EJ, Thomashow MF, Chen TH, Hayes PM (2005) Structural, functional, and phylogenetic characterization of a large CBF gene family in barley. *Plant Mol Biol* 59:533–551
- Skinner JS, Szücs P, von Zitzewitz J, Marquez-Cedillo L, Filichkin T, Stockinger EJ, Thomashow MF, Chen THH, Hayes PM (2006) Mapping of barley homologs to genes that regulate low temperature tolerance in *Arabidopsis*. *Theor Appl Genet* 112:832–842
- Stockinger EJ, Gilmour SJ, Thomashow MF (1997) *Arabidopsis thaliana* CBF1 encodes an AP2 domain-containing transcriptional activator that binds to the C-repeat/DRE, a cis-acting DNA regulatory element that stimulates transcription in response to low temperature and water deficit. *Proc Natl Acad Sci USA* 94:1035–1040
- Tang M, Lü S, Jing Y, Zhou X, Sun J, Shen S (2005) Isolation and identification of a cold-inducible gene encoding a putative DRE-binding transcription factor from *Festuca arundinacea*. *Plant Physiol Biochem* 43:233–239
- Thomashow MF (1999) Plant cold acclimation: freezing tolerance genes and regulatory mechanisms. *Annu Rev Plant Physiol Plant Mol Biol* 50:571–599
- Vágújfalvi A, Aprile A, Miller A, Dubcovsky J, Delugu G, Galiba G, Cattivelli L (2005) The expression of several *Cbf* genes at

- the *Fr-A2* locus is linked to frost resistance in wheat. *Mol Genet Genomics* 274:506–514
- Van Buskirk HA, Thomashow MF (2006) *Arabidopsis* transcription factors regulating cold acclimation. *Physiol Plant* 126:72–80
- Wang Z, Triezenberg SJ, Thomashow MF, Stockinger EJ (2005) Multiple hydrophobic motifs in *Arabidopsis* CBF1 COOH-terminus provide functional redundancy in trans-activation. *Plant Mol Biol* 58:543–559
- Wei G, Pan Y, Lei J, Zhu Y-X (2005) Molecular cloning, phylogenetic analysis, expressional profiling and in vitro studies of TINY2 from *Arabidopsis thaliana*. *J Biochem Mol Biol* 38:440–446
- Xiong Y, Fei S-Z (2006) Functional and phylogenetic analysis of a DREB/CBF-like gene in perennial ryegrass (*Lolium perenne* L.). *Planta* 224:878–888
- Xue G-P (2002) Characterization of the DNA-binding profile of HvCBF1 using an enzymatic method for rapid, quantitative and high-throughput analysis of the DNA binding activity. *Nucleic Acids Res* 30:e77
- Xue G-P (2003) The DNA-binding activity of an AP2 transcriptional activator HvCBF2 involved in regulation of low temperature responsive genes in barley is modulated by temperature. *Plant J* 33:373–383
- Zhao T-J, Sun S, Liu Y, Liu J-M, Liu Q, Yan Y-B, Zhou H-M (2006) Regulating the drought-responsive element (DRE)-mediated signaling pathway by synergic functions of trans-active and trans-inactive DRE binding factors in *Brassica napus*. *J Biol Chem* 281:10752–10759